

## SEARCH REQUEST FORM

Requestor's Name: \_\_\_\_\_ Serial Number: \_\_\_\_\_  
Date: \_\_\_\_\_ Phone: \_\_\_\_\_ Art Unit: \_\_\_\_\_

### Search Topic:

Please write a detailed statement of search topic. Describe specifically as possible the subject matter to be searched. Define any terms that may have a special meaning. Give examples or relevant citations, authors, keywords, etc., if known. For sequences, please attach a copy of the sequence. You may include a copy of the broadest and/or most relevant claim(s).

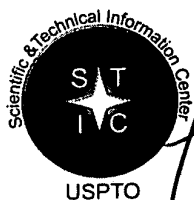
### STAFF USE ONLY

Date completed: 07-08-03  
Searcher: Beverly 24994  
Terminal time: 20  
Elapsed time: \_\_\_\_\_  
CPU time: \_\_\_\_\_  
Total time: 25  
Number of Searches: \_\_\_\_\_  
Number of Databases: 1

Search Site  
\_\_\_\_ STIC  
\_\_\_\_ CM-1  
\_\_\_\_ Pre-S  
Type of Search  
\_\_\_\_ N.A. Sequence  
\_\_\_\_ A.A. Sequence  
\_\_\_\_ Structure  
\_\_\_\_ Bibliographic

Vendors  
\_\_\_\_ IG  
\_\_\_\_ STN  
\_\_\_\_ Dialog  
\_\_\_\_ APS  
\_\_\_\_ Geninfo  
\_\_\_\_ SDC  
\_\_\_\_ DARC/Questel  
\_\_\_\_ Other CGN





# **STIC Search Report**

## **Biotech-Chem Library**

STIC Database Tracking Number: 97680

To: **Phuong Bui**  
Location: **CM1-9E12**  
Art Unit: **1638**  
Tuesday, July 08, 2003

Case Serial Number: 09/938294

From: **Beverly Shears**  
Location: **Biotech-Chem Library**  
**CM1-1E05**  
Phone: **308-4994**

[beverly.shears@uspto.gov](mailto:beverly.shears@uspto.gov)

### **Search Notes**





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 18:49:00 | Search time 2990 Seconds  
(without alignments)

16566.213 Million cell updates/sec

Title: US-09-938-294-2

Sequence: 1 atccgcctcgagttgacc.....gagagattcaatcttagt 1702

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database:

```
Genbank1.*
1: gb_da.*
2: gb_hg.*
3: gb_in.*
4: gb_cm.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pt.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_da.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_hgt_hum.*
31: em_hgt_inv.*
32: em_hgt_other.*
33: em_hgt_mus.*
34: em_hgt_pla.*
35: em_hgt_rtd.*
36: em_hgt_mam.*
37: em_hgt_vrt.*
38: em_sy.*
39: em_hgt_hum.*
40: em_hgt_mus.*
41: em_hgt_other.*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	* Match Length	ID	Description
C 1	282.6	16.6 90341	8 AC024609	AC024609 Arabidops
C 2	282.6	16.6 119942	8 AC000797	AC000797 Arabidops
C 3	253.4	14.9 100685	8 AC023754	AC023754 Arabidops
C 4	224.8	13.2 146081	8 AP001366	AP001366 Oryza sat
C 5	89.4	5.3 2557	8 AY091763	AY091763 Arabidops
C 6	87.8	5.2 75948	8 AC037424	AC037424 Arabidops
C 7	87.8	5.2 117737	8 F6D8	AC023010 Homo sapi
C 8	65.8	3.9 147292	2 AC023010	AC023010 Homo sapi
C 9	64.4	3.8 104486	2 AC109537	AC109537 Neurospor
C 10	58.2	3.4 42273	8 NCBI2N19	AL669988 Neurospor
C 11	58.2	3.4 122151	8 NCBI2H20	AC126306 Rattus no
C 12	57	3.3 152607	2 AC126306	AB026653 Arabidops
C 13	56.4	3.3 35762	8 AB026653	AF128457 Oryza sat
C 14	56	3.3 70311	8 AF128457	AF128457 Oryza sat
C 15	56	3.3 77605	8 AF119222	AF119222 Oryza sat
C 16	56	3.3 142852	8 AF161269	AF161269 Homo sapi
C 17	55.6	3.3 48144	8 AC069166	AC069166 Homo sapi
C 18	55.4	3.3 151771	2 AC128127	AC128127 Rattus no
C 19	53	3.1 23990	1 SC6B10	AL109661 Streptomy
C 20	53	3.1 55493	2 AC096794	AC096794 Rattus no
C 21	53	3.1 141025	2 AC123528	AC123528 Oryza sat
C 22	52.4	3.1 176399	2 AC127769	AC127769 Rattus no
C 23	52	3.1 956	11 PM2D128	AL684743 Penicilliu
C 24	52	3.1 141307	8 AC084763	AC084763 Oryza sat
C 25	51.8	3.0 159184	2 AC130633	AC130633 Rattus no
C 26	51.2	3.0 114741	2 AC111484	AC111484 Rattus no
C 27	51	3.0 136917	2 AP005097	AP005097 Oryza sat
C 28	50.8	3.0 79734	2 AC096238	AC096238 Rattus no
C 29	50.8	3.0 125888	2 AC128139	AC128139 Rattus no
C 30	50.8	3.0 128049	2 AC111232	AC111232 Rattus no
C 31	50.6	3.0 1560	14 HS21E	M29384 Herpes simp
C 32	50.6	3.0 87500	8 AB004308	AB004308 Oryza sat
C 33	50.6	3.0 93419	2 AC121747	AC121747 Rattus no
C 34	50.6	3.0 159563	2 AC112892	AC112892 Rattus no
C 35	50.6	3.0 162473	2 AC127748	AC127748 Rattus no
C 36	50.6	3.0 166421	2 AC098016	AC098016 Rattus no
C 37	50.4	3.0 1285	14 AF202064	AF202064 Bovine he
C 38	50.4	3.0 125020	9 AF429315	AF429315 Homo sapi
C 39	50.4	3.0 134374	2 AC121465	AC121465 Rattus no
C 40	50.4	3.0 144404	8 AB003210	AB003210 Oryza sat
C 41	50.4	3.0 188764	8 AC087726	AC087726 Chlamydom
C 42	50.2	2.9 168532	2 AC121451	AC121451 Rattus no
C 43	50	2.9 2443	9 HM08RASSK2	M30539 Human SK2 C
C 44	50	2.9 10197	1 AE008327	AE008327 Agrobacte
C 45	50	2.9 12006	1 AE009285	AE009285 Agrobacte

## ALIGNMENTS

RESULT 1  
AC024609/c 90341 bp DNA linear PLN 12-SEP-2000  
LOCUS Arabidopsis thaliana chromosome I BAC F14P1 genomic sequence,  
DEFINITION complete sequence.  
ACCESSION AC024609.2 GI:7212002  
VERSION AC024609.2  
KEYWORDS HTG.  
SOURCE Arabidopsis thaliana.  
ORGANISM Arabidopsis thaliana.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
REFERENCE  
1 (bases 1 to 90341)  
Federspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,

[illegible]

STVTPSPYNSKPFEDVGVYNNSTYRENPAPFAVKAVEESSYSDOEVEWMSKSR  
RTSRSEKMTBEMVEYKTEREYTKTMSRSGSDGSSGEFRRTIOTFVSOKOM  
HIRASVNDOLFNQYIRASGDVDPVHWEHGAIPOMNDVDPOLONGVHWMONG  
SVSOLQNSDVPOMONGYILQJLQNDVDPOMOTGSVQLONGVHPONGIYOMKSRBDG  
SGRRHRHGRSRKERO  
/gene="F14P1.6"  
/db\_xref="GI:979584"  
/translation="MWAVEGRICERSKWTGFCGNTGCDSDCKMERASHGACHQ  
PPGAPCYENC"  
complement (81552, .83365)  
/gene="F14P1.7"  
complement (join(81552, .81800, 81920, .82031, 82136, .82233,  
82329, .82436, 82543, .82581, 82670, .82741, 82820, .83068,  
83225, .83365))  
/gene="F14P1.7"  
/note="Unknown protein"  
/codon\_start=1  
/protein\_id="AAF98405.1"  
/db\_xref="GI:9795587"  
/translation="MWAEALPKCEAPLVGLQPAALIDNVAPYDWSLLDIPGRDG  
SLAVQKDELFHMLKELDAHSVAPLKKMGAGSVYTVVGLSVGVATGIGIAYDDK  
OGOLFVSNMGFSQSVISLRRKKGSTOCYVDSGNRTMRPLSSAVKIDADLSK  
EDFGSKMLVIRYAVINLOVQOAIIRPAKEGHSYSDLSAFEMVRSKSLRLES  
GNIDLCPANDEABEHLREBOGPEALHFFRGRNAVTVLSEKCIACKIKDKEVTH  
ISAIGETVATDAGADLFAAGFLYGLIKGLSLERCCVSGSGSVIRALGDEVTE  
NMQWHLQQLKGLPVPDIHN"  
complement (85358, .87513)  
/gene="F14P1.8"  
complement (join(85358, .85648, 86056, .86207, 86310, .86435,  
87001, .87144, 87399, .87513))  
/note="Unknown protein"  
/codon\_start=1  
/protein\_id="AAF98404.1"  
Query Match 16.6%; Score 282.6; DB 8; Length 90341;  
Best Local Similarity 57.9%; Pred. No. 2,46-50;  
Matches 702; Conservative 0; Mismatches 304; Indels 207; Gaps 3;  
QY 469 TCGAGGTTTACACGCTAGACGACAGAGGCACTGATTCCTTAACCTATCTG 528  
DB 46971 TTCAAGGTATACGACAAAGTCAGAAAGCTATAGATACGCTCTCAAGTCTGATCG 46912  
QY 529 TTATCTTAAACACTGCTGTGTGCGACAGTGGCTTGACCTGTTCTGAAGATCATGTC 588  
DB 46911 TTGTTTAAACAGCTGCTGTGTGCGAAGTGGCTTGATGACGTTCTCAAGGACATGTC 46852  
QY 589 CTTAAAGTCTCCGAAAGTTTGTGTGAGTCACATGAAGAGCGTGGCACTTAAAG 648  
DB 46851 CTTAAAGTCTCTTAAGTACTCTGAGATTCATGAGAGAGAGTCTATTTTCAAC 46792  
QY 649 TTGAATATGCTCAACATCTCCCTTGTGTGAGACCATGATGATTTCTCATACACGG 708  
DB 46791 CGGATTTAGTCAAGACCTTCGTTGTGACAGGTGATGATGATTCGACCAACAG 46732  
QY 709 CTGAGTATGGAATAGCAGACCTAGGATCGC----- 740  
DB 46731 CTGATATCTGAAAGACAGAACTCATGATCGCTTAGGGTCTGTTTCTCTCTAGAGTC 46672  
QY 741 ----- 740  
DB 46671 TTGGATATATTAGTTTACGTTATGTCATTAAGCTTAACAGTGTATTCTTTATTGTT 46612  
QY 741 -----CTGAAATATAGATGCCACAACTTATGTTTACCTGCGGAGATATGTAAGACT 796  
DB 46611 GTTACACAGATTTAAATGCTTAAACTTATGTCGTGCACTTGAAGAAATAGCAAGGCTT 46552

QY 797 AATGAGATTGCTGAAGACAATGTGCGAAGAAAGTCTTACGGAAATATTCGTGTAATC 856  
DB 46551 GATGAAAGTATGCTGAAGATATGTTTCGCAAGAAATGTTCTCGTAGACCAAGTCCGAGATC 46492  
QY 857 CTTGAGATGAGATGAGATCTCCGTTTGCAATTAATAAC----- 900  
DB 46491 TCTTGAGATGAGATGAAGACATATTTGTCATTTAATATGTTAGTCAATTTCTTAC 46432  
QY 901 ----- 900  
DB 46431 ATTCTTCTCTTCTTATGTTCTTTAAGATCATATCTGCAAAAGACTCAAAATTTTGG 46372  
QY 901 -----GTGATACAGGAGAAAGGACAG 924  
DB 46371 TTAATTTGCTTTGCTGACTGCTGCTCTTTAGCAGAGTGATCTCGAGAAAGGCCCAAG 46312  
QY 925 ACTTATTTCTTCAAGCATTTTATCAAGGCTTTGACCTCATCAACAG---AGAAACTTA 981  
DB 46311 ATCTATTTCTCCAGGCTTCCATATGAAGTCTTAAAGTATCAAGAGACTTAAAGAACTTG 46252  
QY 982 AAGTCCCTAGATATCATGCTGATGTTGTGGAAAGTATGTTAATGCTCAACCAATTTG 1041  
DB 46251 AGCTTACCAATCATGCTGATGTTAGTAGAAGCGATATAGCGCACAGAGAAATTTG 46192  
QY 1042 AGACTCAGTTACGTGACTTTGTGTGAAGACACGATTCATGACCGTGTCCATTTTGTGA 1101  
DB 46191 AGACGAGCTACCAACTTTGTCCAGAGATGAACCTTCAAAAATTTGTCACTTTGTCA 46132  
QY 1102 ACAAGACATTGGCAGAGGCGCCCTTACTTGGCAGCAATGATGCTGTTCAAGAAATCTC 1161  
DB 46131 ACAAAACATGAAGTAGAGACCAATTTTAGCAGCTATTAGAGTTCTGTGCCAAATCTCC 46072  
QY 1162 AGGCGCTGAGAAATCTTTGGAAGATTAACATTAAGCAATGCGCTTAACTTCCAG 1221  
DB 46071 AAGCCAGAGAGAAATCTTTGGAAGATTAACATTAAGCAATGCGCTTAACTTCCAG 46012  
QY 1222 TTTTGGGACGCTGTGTGAGAGGACCAAGAGATCGTCTGGAAGGCTGCACTGCGCTTC 1281  
DB 46011 TACTTGGGACGCTGAGAGGAGGACCAATGAGATGTGATGACAGAAACAGGCTAT 45952  
QY 1282 TGCATCTGCTGGAAAGAGGCGCTGCGCTTTGCAAAAGACATGTGATCGCA 1341  
DB 45951 TGCACACACCTGTGAAGATGCGCTGTACTCTTCCCAAAAATTTGTAACCTGCGCA 45892  
QY 1342 GCCACGCCAGACAGAGGCTCTCATGAGGAGAAAGGCTATGCGAGGATGAGAAATGT 1401  
DB 45891 CTTAGCTGAAGATGAGAAATTAATGAGGAGAAAGGATATGAGAGGATTAAGATGT 45832  
QY 1402 TCATGAGACACACATGCTGAGAGATCGCGCGGTGTTGAAGATGCTCCGAGAAAT 1461  
DB 45831 TCTTGAACATCATATGTCTCATGAATAGCTTGTGCTCAGAAAGTGTTCACACATG 45772  
QY 1462 CACAGAGCACTC 1474  
DB 45771 CAAAAATTCACCTC 45759  
RESULT 2  
AC007797/c 119942 bp DNA linear PLN 12-SEP-2000  
LOCUS Arabidopsis thaliana chromosome I BAC F6F9 genomic sequence,  
DEFINITION complete sequence.  
AC007797 AC007797  
AC007797.7 GI:7839909  
KEYWORDS HMG;  
SOURCE Arabidopsis thaliana.  
ORGANISM Arabidopsis thaliana.  
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
REFERENCE 1 (bases 1 to 119942)  
Fedorispiet,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,

Altafi, H., Nguyen, M., Lam, B., Southwick, A., Miranda, M., Brooks, S., Buehler, E., Chao, Q., Chin, C., Chou, J., Choi, E., Gonzalez, A., Howng, B., Johnson-Hopson, C., Khan, S., Kim, C., Koo, T., Lee, J.H., Lenz, C., Liu, A., Liu, S., Mukharrey, N., Pham, P., Sakano, H., Shin, P., Tortum, M., Vaynsberg, M., Yu, G., Becker, J., Theologis, A. and Davis, R.W.  
Unpublished

JOURNAL  
REFERENCE  
AUTHORS  
2 (bases 1 to 119942)  
Federispiet, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Araujo, R., Huizar, L., Rowley, D., Buehler, E., Dunn, P., Gonzalez, A., Kremeneckaja, I., Kim, C., Lenz, C., Li, J., Liu, S., Lueros, S., Schwartz, J., Shin, P., Tortum, M., Wysockaja, V., Walker, M., Yu, G., Becker, J., Theologis, A. and Davis, R.W.  
Direct Submission  
Submitted (12-JUN-1999) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

TITLE  
JOURNAL  
3 (bases 1 to 119942)  
Federispiet, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Nguyen, M., Lam, B., Southwick, A., Bel, O., Buehler, E., Chin, C., Chou, J., Choi, E., Dunn, P., Gonzalez, A., Howng, B., Kim, C., Koo, T., Lee, J.M., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharrey, N., Pham, P., Sakano, H., Schwartz, J., Shin, P., Thaveri, A., Tortum, M., Vaynsberg, M., Walker, M., Yu, G., Becker, J., Theologis, A. and Davis, R.W.  
Direct Submission  
Submitted (16-MAY-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

TITLE  
JOURNAL  
4 (bases 1 to 119942)  
Federispiet, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Nguyen, M., Lam, B., Southwick, A., Becker, J., Theologis, A. and Davis, R.W.  
Direct Submission  
Submitted (12-SEP-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

COMMENT  
On May 16, 2000 this sequence version replaced gi:7547094. Bases 60,267-119,942 of BAC clone F6P9 overlapped with bases 1-59,676 of IGF clone F4P1. AC024609 and bases 1-3,490 of BAC clone F6P9 overlap with bases 89,221-92,710 of BAC clone T20H2. AC022472.

FEATURES  
source  
Location/Qualifiers  
1. 119942  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/chromosome="1"  
/clone="F6P9"  
2448. 4393  
/gene="F6P9.1"  
join(2448..2640,2712..2995,3068..3502,3589..3727,3808..3974,4064..4393)  
/gene="F6P9.1"  
/note="Similar to endo-beta-1,4-glucanase"  
/codon\_start=1  
/protein\_id="AA012562.1"  
/db\_xref="GI:10086502"  
/translation="MVAKPRSPCCSCVETGUTLLATITVITPHRNSHDDGGSNYKYNANLKTAMQFIDKSKLNNEISNRGSGLKDGSEASIDLSKLYADAGHMKEGPMATATATVLSWSTLIEGDQMASLMDLHAKSLDKWTDFLINHBPBNLYIOVGPDVTDHCDRPEPTWTRKRLTKLDTKPTETVAETAAAMAASLVFKESPSTKXSS TLLKTKAQLEDFADNNRGSYVNIPEVGSYSTGYDELLMAASMYHAETEOYALD FVSENGEFGFGSPSMFSWMDKLPGTHILSLRTFFKGLSGSGKLGQGEKELAEVW CGLLPSSTATASRTDGLIIVSEWNALQHPVSAFLATLYSDVMTLGVKELSCSDS SKPSPDLKFKARSQADYMLGKVPKEMSYLVGGEKYPEFVHHRGASLPADATTCGDCG FKMLNSDEPNVAVGALVGPFLNDFIDARRNNSQNEPSTYSALVGLSLSVTTSSVSFSF"

gene  
CDS  
/gene="F6P9.2"  
join(6059..6600,6686..7091,7183..7443,7547..7633,7747..7881)  
/gene="F6P9.2"  
/codon\_start=1  
/product="sulfate adenylyltransferase"

/protein\_id="AA012541.1"  
/db\_xref="GI:10086481"  
/translation="TMSLMRSSVSHITLQPNRSPSSFTQISFLSSNNPFLNLYKRNLTQWSKWTYKSLIDPDGELVELIYETIEIGVKAASEMPKYKNOIDL EWHVYISGMAFPLKGMREDEYLOSHFNRLKNGTVPNNSLPYIALDDDTKCOI GSEENVALVCPQDIIIGSLRSEYIKKHKEERLARTWGTSSPGLPYEVETRSGWL IGGDLLEPPIKYNIDLDHYRLSPKQLEPDRQADNPAQLENPHNGALLND TIRRLLEGVKNPVLILHPLGCGFTKADVPIDRMQCHSKVLDGQVLPKTTVTFP SPNHYAGTEVQWMAKRNINAGANTYIVGRPPMGHPTEKRDLYDPHGKRLVSNAP GSEKNTLIPFKNVAVDITERRKAFDPSPRAKEPLFTSGTKMTARTAGENPDGFCP SGNNLVVYTESIQSEBARQOAVSA"  
complement(8394..9399)  
/gene="F6P9.3"  
complement(join(8394..8526,8930..9215,9921..9399))  
/gene="F6P9.4"  
/codon\_start=1  
/product="vacuolar H<sup>+</sup>-pumping ATPase"  
/protein\_id="AA012542.1"  
/db\_xref="GI:10086482"  
/translation="MASTSGDETAPFEGFLGAALVPSGCAAGYATKAGVAVSM GVMRELVMSIVPMAGVLIGLIIAVISTGINPAKSYLPDGYHLSGLAC GLAGLSAGMAIGIVGDAGVRANAQPKLFVGMILLIFALALAYGLIVGILLSSRAG QSRAR"  
14164. 15678  
/gene="F6P9.3"  
14164. 15678  
/gene="F6P9.4"  
/note="Unknown Protein"  
/codon\_start=1  
/protein\_id="AA012543.1"  
/db\_xref="GI:10086483"  
/translation="MEMOLHNDRVWYDRTPFGSNTSLPNGCRNPQDAVSKTID TASHIEDVANTIRPLTVQSNWSSGSRPQGVNQTGGDRGELKTRTSPNNOC CDWEMNNGLKKRWYASNNHLPDGOIKVGGCCGFRVEFPPTTNNVALPLAE THDQSGNNNXPVPMNTDGLPMFANRRILLIDVYKATVKTFAIPAGSDRNPST GSAVILLPLKULENDVETEVLCGAPKGSYNLARKKIFVALDTCRKINDAPREW AVENKPFARVNGDITPLRNGVLLINGSGSTGLAMELCRTVALDPLDIEKPNVSRF ESKRPITTPMWHSAAILIRGRVAVLGGSNPAHFVLEPFLSELFKSPVLSOR FESNLRKIIISPPEQSMIKYGTNLKLFSTGSGVTTPAKVMTVPPTTWSFAMNQVR LVLDNVFTRKSGSPMYEVOVTRPSANLAWGYVMIIVVNDIIPSGVVKLE"  
complement(16924..17682)  
/gene="F6P9.5"  
complement(join(16924..17094,17228..17318,17417..17496,17611..17682))  
/gene="F6P9.5"  
/codon\_start=1  
/product="Putative histone H3"  
/protein\_id="AA012563.1"  
/db\_xref="GI:10086503"  
/translation="MARTKOTARKSGTGKPKRELATKAARKTRPRGCVRAHRR PGVALRETRKOKSTDLLIRKLPFORIVRELAADFVKDLRFQSHAVLALQEAAYLV VGLFEDNLCATIHAKRVIMKQIDLRIRGERA"  
18998. 22068  
/gene="F6P9.7"  
join(18998..19093,19182..19243,19336..19403,19510..19631,19701..19824,19937..19968,20276..20393,20887..21016,21098..21220,21301..21354,21480..21661,21734..22068)  
/gene="F6P9.7"  
/note="Unknown Protein"  
/codon\_start=1  
/protein\_id="AA012544.1"  
/db\_xref="GI:10086484"  
/translation="MAEAMNSSEKKEDESEKGBELLFCGATAMDIIIGRKAGENIL VSPFLRLPVGAVNIRFVATGCAFHCVALDVEGRVGNENKGGUGHDMDIORDPT VVSGLSKRIIVAAAGRNHTVVASDDGSLDFGNMKYGGOLGSLAKNGFVSVESPT LPCVSDDEVTVNACGADFTWVLSSTEGASILLAGIIPQYGLFGTGNENFKNSSVRL AYEAGRPRAIASLAGETIVKACGTNHTFWSYIKRFLPLSTGGGLQIMWKIKTN NGDDMYPKPMWDLGNNLIRMDSGSMHFFVAGASQISMGHGOYERLCEYGNQKSL AAPKVDMLGEMHWNGTDIADRLBOYTKRAREPFRALNKRIYISRSRFTSSSYSL HCLICOLEVNDGKSLAEASVEVEKERTLAPKQAVTKRGASRKAKASSSBED SDENSDKEKEVQSDSDSDSDSEGEANGKQASRGRGRGARGRGRGRTSSNGKAPP VKTGRRGKPRKS"  
24177..27256

gene

CDs /gene="F6F9.8"  
join(24177..24269,24579..25172,25331..25510,25592..25651,  
25766..27256)  
/gene="F6F9.8"

/note="Unknown Protein"  
/codon\_start=1  
/protein\_id="AAG12559.1"  
/db\_xref="GI:10086499"

/translation="MGSPASSSCRLISCGDDTSDADPNSTALENKSSGDKRMSFR  
KSGKORGLITVSEETTPASRETTLESALKSPSPNNVSEKQOSFVDEKSKQ  
LPVYTAAPVDEKTSQVIEKTELLSYERDIDHREVPVIESEKREYDDELIG  
TELQSGADADAKIEDTSEVEMASKVPESSSDVITIRSEKREYDDELISYI  
VVIQAAVGLARRELSKSVIKLQAVRGHLVNSQMSGLCVALYKQAMVRRAR  
HSTKDSRVASATSDSEPNAAQKLENFPAHLMESTPKFINKCDPTKESANN  
WLERMSVPRPKTSKANLITTEBONLSEYONVDSIPQADPVNSDITVENKPTMPSY  
EASVEQONVELSETEKMSQYDSPEASAEVYDISOQPLAARGDSILDEPEYVDQI  
KSLKRVSNPSFIAOSKFELITSTGSKNMTLSKDVLADEGKTDI0SPDTYGT  
IKHSLIEDVTPALSGSHCELTSLVTSIDLTLEKSGAEGRPREVAKLIEDDFKTD  
QALITIDYKDATSMGTEDPEKKEVKNADVEI8ATHEPVI5TFDGSKKRAEDESG  
PQAYALSEGLTPMTTTSQTPASQSSSVARBRKESGSGSQRKXSKKITSBK  
QELTGATBQEBGKQSGRISFGTDQABESSGQNSLPRPNQPTQSAKAYQSH  
NSPSSPDLOERDVVASAKRHSILPGVTNGQVSPRIQSA5Q0QGTGKNNLLDSEI  
CNHSSQP"

gene 28911..31062  
/gene="F6F9.9"  
join(28911..28985,29480..29677,29976..30100,30198..31062)  
CDs /gene="F6F9.9"

/note="Hypothetical Protein"

gene /codon\_start=1  
/protein\_id="AAG12545.1"  
/db\_xref="GI:10086485"  
/translation="MQENAPRTMSDVORVLAISNKFDMVLFISEDSPOVSQSESOD  
HLOAKSLASHPSDDILPGFCGFSANESQIKLSDIPLIKMCSVOLILDEMWVVA  
GDEKSEVQONERELELVLEAFYPGASIPENPSVPADESDHDDQITVILILEVD  
DDIMDASDPFQSGVDGTPESITDENTSTLPGPDIMALSAISNKEGSM  
IDDLIKLISNPKLENIYANRGSAGSVNSSTLSYSSVTVTAAPISNQG  
FYAOPITTHIPMAATYPRARODOPNVGAPAPRABSVYNNLLOHGBRDEPPTOHLG  
YRNLQPGGAPNPEMNSNNQKPRDSKXIKMKACMFNBARCRRGANCMYQHDAT  
PYQPNLNNGININTSDMQAKRMRPDK"

complement(31753..35330)  
/gene="F6F9.10"

Query Match 16.6% Score 282.6; DB 8; Length 119942;  
Best local Similarity 57.9%; Pred. No. 2,56-50;  
Matches 702; Conservative 0; Mismatches 304; Indels 207; Gaps 3;

QY	469	TGCAAGTTTACCACTAGAGACAGAGGAGTTGATTTGCTTAAAGCTATTTGG	528
DB	107237	TTCAAGGTATACAGCAAGGACGAAAGCTATAGATACAGCTCTCAAGTCTGATCTG3	107178
QY	529	TTATCTTAAACGCTGCTGTGCGAAGTGGCTTGACCTGTTCTGAAAGATCATGTC	588
DB	107177	TTGTTTAAACCTGCTGCTGTGCGAAGTGGCTTGATGAGATTCCTCAAGGACAAATTC	107118
QY	589	CTAAGTCTTCGGAAGATTTTGTGTGATCATGAATGCGTGGGCTTACTTTAAG	648
DB	107117	CTAAGTCTTCCTTAAGTACTCTGAGTGAATTCAGAGATGAGAGGCTCACTAATTTCAAC	107058
QY	649	TTGATATGTCAACATCTTCCCTTGTGTGAGCATGATGATTTCTCATCAACG	708
DB	107057	CGATTTAGTCAAGCACTTCCGTTTGTGAGGAGCTATGATTTGACACGCAACAG	106998
QY	709	CTGAGTATGGAATAGCAGACTAGGATCGC-----	740
DB	106997	CTGAATACGGAAGAACACTCAGATGCTTAGGGTCTGTTTCTCTCTGAGAGTC	106938
QY	741	-----	740
DB	106937	TTGCAATATATAGTTTACGTTATGTCAATAGCTTAAAGTGTTTATTTCTTTATGTT	106878
QY	741	-----CTGAAATATGAGTCAACAACTATGTGTTTACCTGGGGATATGTAAGAACT	796
DB	106877	GTAACACAGGATTAATAAGCTTAACCTATGTCGTCACCTAGGAAATAGCAGAGCTT	106818

QY	797	AATGAGATTGTCTGAGAGCAATGTGCAAGAAAGATCTTACGGGAAATATTTGTAATC	856
DB	106817	GATGAAAGTATGCTGAAGATAGTTTCCCAAGAAATGTTCTCCGAGACAGTTGAGATTC	106758
QY	857	CTTGAAGTACAGGATGAGATCTCCCTGTTTGCAATATAAACA-----	900
DB	106757	TCCTTGAGGAGAAATGAGACATCACTATTTGGCATTTTAATAGTTAGCATTTCTTAC	106638
QY	901	-----	900
DB	106697	ATCTTCTCTCTCTAGTTCTTTAAGATCATATCTGCAAAAGACTCAAAATTTTGG	106638
QY	901	-----GTTATACAGGAGAAAGGACAAG	924
DB	106637	TTAATTTGCTTTGACATGAGTCTTCTTTATGACAGGTGATCTCGAGGAAAGGCGCAAG	106578
QY	925	ACTTATTTCTTCAAGCATTTTATCAGGCTTTGACCTTCATCAACAG---AGAACTTA	981
DB	106577	ATCTATTTCTCCGAGCGTTCCATGAAAGTCTTAAAGTATCAAGAGACTAAGAAATTTG	106518
QY	982	AAGTCTTGAATATACATGCTGATTTGTGGAAAGTATGTTAATGCTCAGACCAATTTG	1041
DB	106517	AGCTTCAACATGATCATGCAATGATGTTAGAGAAAGCATATAGCGCACAGAGAAATTTG	106438
QY	1042	AGACTGATTTAGCTGATCTTTGTGTGAAAGAACATTCATGACCGGTTCATTTTGTGA	1101
DB	106457	AGACAGAGCTACGCAACTTTGTCCAGAGATGAACTTCAAAATTTGTCACCTTTGCA	106398
QY	1102	ACAGACATTTGGCAGAGGCGCCCTTACTTGGCAGCAATTGATGTCTTGTCAAGAAATTC	1161
DB	106397	ACAAAACATGAAAGTATGAGACCAATTTTACAGCATTTGAGAGTTGTCGCAAACTCC	106338
QY	1162	AGGCGCGTGAAGATCTTTGGAAGATTAACAATTTGAAGCAATGCAATTTAGTTCCAG	1221
DB	106337	AGCCAGAGGAGATCTTTGGAAGATTAACAATTTGAAGCAATTTAGTTAGCTTCCG	106278
QY	1222	TATTTGGGACGCTGCTGTGAGAGGACCAAGAGATCGTCTGAGCGCTGACGTGCTTC	1281
DB	106277	TACTTTGGGACGCTGAGAGAGGACCAATTTGATGTAAGCAACAGGCTCAT	106218
QY	1282	TGCATCTGCTGGAGAGAGGCGCTGCTTGTTCGAAAGAAATGCTGAGCTCGCA	1341
DB	106217	TGCACACATGCTGTAAGATGCGCTGTATCTTCTGCAAAACATTTGTAAGTGGCGA	106158
QY	1342	GCCAGCGGACAGAGGCTCTCATGAGGAGAAAGGCTATGAGAGGATGTAAGAAATGT	1401
DB	106157	CTAAGTGAAGATGAGAGATTAACAATGAGGAGAAAGGATGAGAGGATTAAGATGT	106098
QY	1402	TCATGAGACACCATGCTGAGAGATCCGCGGTGTTGAAGATGCTCGAGAAAT	1461
DB	106097	TCCTTGAACATCATATGCTCATCATGATAGCTTGTGTCTCAGAGAGTGTTCACATG	106038
QY	1462	CACAGAGCACTC 1474	
DB	106037	CAAAAATTCATCTC 106025	
RESULT 3			
AC023754/c 100685 bp DNA linear PLN 13-SEP-2000			
LOCUS Arabidopsis thaliana chromosome I BAC F1B6 genomic sequence,			
DEFINITION complete sequence.			
ACCESSION AC023754			
KEYWORDS AC023754.3 GI:7212003			
SOURCE HMG			
ORGANISM Arabidopsis thaliana			
Arabidopsis thaliana			
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.			
REFERENCE 1 (bases 1 to 100685)			
AUTHORS Pederspiel,N.A., Palm,C.J., Conway,A.B., Conn,L., Hansen,N.F.,			

Altafi, H., Nguyen, M., Lam, B., Southwick, A., Miranda, M., Brooks, S., Buehler, E., Chao, Q., Chin, C., Chlou, J., Choi, E., Gonzalez, A., Homg, B., Johnson-Hopson, C., Khan, S., Kim, C., Koo, T., Lee, J.M., Lenz, C., Liu, A., Liu, S., Mukharkey, N., Pham, P., Sakano, H., Shin, P., Toriumi, M., Vaysberg, M., Yu, G., Becker, D., Theologis, A. and Davis, R.W.

JOURNAL  
REFERENCE  
AUTHORS  
2 (bases 1 to 100685)  
Unpublished

Altafi, H., Nguyen, M., Lam, B., Southwick, A., Bel, O., Buehler, E., Chlou, J., Choi, E., Dunn, P., Gonzalez, A., Homg, B., Kim, C., Koo, T., Lee, J.M., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharkey, N., Pham, P., Sakano, H., Schwartz, J., Shim, P., Thavert, A., Toriumi, M., Vaysberg, M., Walker, M., Yu, G., Becker, J., Theologis, A. and Davis, R.W.

JOURNAL  
TITLE  
Direct Submission  
Submitted (17-FEB-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

3 (bases 1 to 100685)  
Federapfel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Nguyen, M., Lam, B., Southwick, A., Bel, O., Buehler, E., Chin, C., Chlou, J., Choi, E., Dunn, P., Gonzalez, A., Homg, B., Kim, C., Koo, T., Lee, J.M., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S., Mukharkey, N., Pham, P., Sakano, H., Schwartz, J., Shim, P., Thavert, A., Toriumi, M., Vaysberg, M., Walker, M., Yu, G., Becker, J., Theologis, A. and Davis, R.W.

JOURNAL  
TITLE  
Direct Submission  
Submitted (09-MAR-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

4 (bases 1 to 100685)  
Federapfel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F., Altafi, H., Nguyen, M., Lam, B., Southwick, A., Becker, D., Theologis, A. and Davis, R.W.

JOURNAL  
TITLE  
Direct Submission  
Submitted (13-SEP-2000) DNA Sequencing and Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304, USA

COMMENT  
On Mar 9, 2000 this sequence version replaced gi:7143418  
Bases 1-32,040 of IGF clone F1B16 overlap with bases 97,508-129,547 of IGF clone F10A5, gb|AC006434.  
e-mail for correspondence: arab@sequence.stanford.edu  
Genes with similarity to proteins in the databases are named 'putative', '-like' or 'similar to'. Genes that have EST similarity but no significant protein similarity are described as 'unknown proteins'. Genes that are annotated based only on gene prediction software are described as 'hypothetical proteins'. The gene prediction programs used to predict genes include: Graft (Informatics Group, Oak Ridge National Laboratory, <http://combio.ornl.gov/section/index.html>), GENSCAN (Chris Burge, <http://genomic.sanger.ac.uk/gf/gf.shtml>), and Necplantene (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/Necplantene.html>).

FEATURES  
source  
1..100685  
Location/Qualifiers  
/organism="Arabidopsis thaliana"  
/db\_xref="taxon:3702"  
/chromosome="1"  
/clone="F1B16"  
1..32040  
/note="overlap with bases 97,508-129,547 of IGF clone F10A5, gb|AC006434. See Genbank record for BAC clone F10A5 for annotation in this region."  
complement (32986..33822)  
/gene="F1B16.1"  
complement (32986..33822)  
/gene="F1B16.1"  
/note="Unknown protein; unknown protein"  
/codon\_start=1  
/protein\_id="AG13069.1"  
/db\_xref="GI:10120444"

gene  
CDS  
1..32040  
/note="overlap with bases 97,508-129,547 of IGF clone F10A5, gb|AC006434. See Genbank record for BAC clone F10A5 for annotation in this region."  
complement (32986..33822)  
/gene="F1B16.1"  
complement (32986..33822)  
/gene="F1B16.1"  
/note="Unknown protein; unknown protein"  
/codon\_start=1  
/protein\_id="AG13069.1"  
/db\_xref="GI:10120444"

gene  
CDS  
43756..46560  
/gene="F1B16.2"  
join(43756..44347,44507..44562,44700..44966,45493..45755,46200..46560)  
/gene="F1B16.2"  
/note="Similar to cytokinin oxidase"  
/codon\_start=1  
/protein\_id="AG13068.1"  
/db\_xref="GI:10120443"  
/translation="MTSSFLITLTPKILKLIANGLVGSELSLIGALIDVGHFTVHP  
SLIASVSDPDMKSPFEPLAVLPSSNAEDVARLVKTAGSATAFPVSARGHSLNG  
QAAAGNVVVENHGVGTGTPKPLRPDEMVLVWGLMDVLPKTLHGLAPKSWT  
DLYLVGGLISNAGISGQAFHGPQISNVLIEDVVTGQFGITIRARISLEPAPORV  
WIRVLVSFVFTEDQEDYLISWGLQKFDVYVDEGLVNNMRSFSPSPNPKVI  
SSVSNGLSVLCLEITKYNHDSSEIYDQEVETLAKKNFIPSPFTDLOVYFLDR  
VHKAELKRSKLMVEVHPHMLNLPVPSKRSIDSGVFGKLGKSGPLIYPMNKD  
KMDERSAVTPDEEVYLVALLRSALDGEETQKLEYIKDNRILRFGCAKATVKKO  
YLPHATQGEVVAHFGDKWDRFRLKAPRPRHILATQGRIFQNPISLTFPSSSSSS  
AASW"

gene  
CDS  
complement (46848..48185)  
/gene="F1B16.3"  
complement (join(46848..46943,47040..47128,47474..47537,47773..47875,47963..48066,48156..48185))  
/gene="F1B16.3"  
/codon\_start=1  
/product="Putative ubiquitin-conjugating enzyme"  
/protein\_id="AG13066.1"  
/db\_xref="GI:10120441"  
/translation="MSSGAPSRKTKLAKTNRLKELVEMKNNPPTGKHKATYDNLQ  
KMLIEVAGKGLVADITYQVDPPEHYPMSPQVITLHAPLHPHILVSNGLICLDI  
LYDSWSPAMTVSSICITSLMSLSSSTEKORPTDNDRYKCKNGKRSKETRWFFHDK  
V"

gene  
CDS  
52291..53501  
/gene="F1B16.4"  
join(52291..52584,52917..53501)  
/gene="F1B16.4"  
/note="Hypothetical protein"  
/codon\_start=1  
/protein\_id="AG13063.1"  
/db\_xref="GI:10120438"  
/translation="MEDPRVHHCSSLRGTLDSRYAAVACVCEVIEDIGREVEYLC  
NNILINDLPGRRRPFPALSEIKSLCSGSPMSJPEHNEHLIKITKLSLQDYER  
PEQYNOLQOVISPEEIAAGSSKYTGALQAMTRPGLSEALISQTNVYRRT  
ISHQVYKTISSGLSGLSLPQNTTSSLSORGLVQGGPQAKPRGIPETSYAIR  
AMLFHPLHRSVPLIKKNTSDBSCLKCMRNISLIVAKTYIRMKRSWCKLKOD  
FRKRTKPTMLILTFP"

gene  
CDS  
complement (54491..56307)  
/gene="F1B16.5"  
complement (join(54491..55074,55197..55351,55498..55769,55872..56014,56253..56307))  
/gene="F1B16.5"  
/note="Hypothetical protein"  
/codon\_start=1  
/protein\_id="AG13070.1"  
/db\_xref="GI:10120445"  
/translation="WGTDFVSVICINSYDSRGLPMLMELALRGVADVITNQ  
KPLEDEVVYSLEHMLDRGVQVTSKQKQAVDLSLADILVNTAVAGVADVITNQ  
NVKVLPKLIMNTHMGHGFNDLVKLPVAAAMDSHATGAKWKRATQAGIKM  
PKTYVHLNLSKELMEVAEDSVAKRLREHRLVIGRNDLLFGINSRGAGGQD  
FLRAFHESLERKEKLOVPTMAVVGSDSKTKEETLRAPREKMETLPPHFPN  
KTLTVAPYLAIDVLYONSQARGEGRITTEBAAFTLPVIGKAGKTEIVNAGTGV  
LHSAKKEGVPIAKNTVKTATVDELRIRMKNGYEVYKEMPLEHMSHRIASVKEV  
LQKARARAR"

gene  
CDS  
55918..61713  
/gene="F1B16.6"  
join(55918..60655,60734..61110,61188..61248,61315..61713)  
/gene="F1B16.6"

```

/note="Similar to homeodomain proteins"
/codon_start=1
/protein_id="ANG13065.1"
/db_xref="GI:10120440"
/translation="WAVYFNSVGMOSLYQSSITLYNQOQQOQOQSSSSAASFSEIYS
GIVRNNEVFIPTSTDAVNGNVTASNDLSFHGGGLSLGNOIQSAVSPFOH
ONLSNQLSYNNLNPSTSDENGKSLSHQHSDIPLFSVYNNNGNNGVGVNRYRE
TSGFVSVLSRSLKPTQQLDEVSVVRDKLGNKKMKNDKODFHNQSSDNIITEDD
KSQSOELSPSEROELQSKSLTWDEVDKRYQYHOMALASFEWYGLGAAE
YTSVALNRIISRHPCLDAIKEOIVIRCKGERETSDQGERIPRLRYDLQRLQOR
ALHOOLGWRPAMPORGLPENSIVILRAMLFEPHRLPYKSEKIMLSTKOTGLSKNO
VANWFINAIVRLMKMTLEMYKEKRGESSEILSNQDITKQKQTSOLKHESSSSNO
QNGNNNNNITPTSDAQNIVFADPEKPRATTGYDLSIMYHGFGLIDNNRYGLGNO
QDGRYSNPQLHDFVY"
/complement(62425..64698)
/gene="F1B16.7"
/complement(join(62425..63429,63528..63559,63863..64074,
64640..64698))
/gene="F1B16.7"
/note="hypothetical protein"
/codon_start=1
/protein_id="ANG13067.1"
/db_xref="GI:10120442"
/translation="MPFIRVSCFDPENLILFWRYADEIBGSPYSPYASRGISM
DKMSLSEKGLTSEGGTPPDGHLGTASQKATPEMSTWVSPSSGLASDLIEV
KNIVDPGLVPSAOKPLFTSPILPNVPSDLISATRLPLPKTSPPRRARSPGHOLF
ROYSDQITGLKSPINNYISISEKSSFLVSTCSNDPATGSOVASSEGSMNTSELY
AYSQRPWSPSEHLGSGRRKLGSSSRFSPSVVOQVCSKRLITRESSIATPE
LPIAVALACGHYHAECELTETTDIKYQAPACTTGTEKRVAKITRKLIVAEARA
KQYRCKNRVDSYGEBCDEFVQKMGREKAKIKELACSSKSSSKNSPLKHPAS
ISSKMKPSSKOSALKGFWRHNNRSSSTLEKTLTNIQS"
/complement(69063..69566)
/gene="F1B16.8"
/complement(69063..69566)
/note="similar to DNA-binding protein"
/codon_start=1
/protein_id="ANG13064.1"
/db_xref="GI:10120439"
/translation="WGSSTGNCSSVSTTGLNAGESBDRQNDLDERKKRQKSR
BSARRSRMKRKHLDLTAYQTHLRKKNQIYAGIAYTQHYVITAEENDILRAQVLE

```

```

Query Match 14.9%; Score 253.4; DB 8; Length 100685;
Best local Similarity 65.4%; Pred. No. 5e-44;
Matches 371; Conservative 0; Mismatches 196; Indels 0; Gaps 0;

OY 901 GTGATTCACGAGAAAGGACAAAGCTTATTTCTTCAAGCATTATATAGGCTTTGCAAC 960
DB 55074 GTGATTCGCGGAAAAGGACAGATCTGTTCTCCGAGCTTCCACGAAAGCTTGAANA 55015
OY 961 TCATCCACACGAGAAACTTAAAGCTAGATATACATGCTAGTGTGGGAAGTGATG 1020
DB 55014 GATCAAGAGAAAGAACTTCAGATACCAACATGATGAGATAGTAGAAGCGACA 54955
OY 1021 TTAATGCTCAGCAATTTGAGACTCAGCTTACGCTTGTGGTGAAGAACACATTC 1080
DB 54954 TGGCAAGACAGCAAAATTCGACAGAGCTACGCAACTTGTGCGAAGAAAAGAACTTG 54895
OY 1081 ATGACCGTGTCCATTTTGTGAACAAGACATTTGGAGTGGCCCTTACTTGGCAGCAATTG 1140
DB 54894 AGAACTTTGTCACTTGTCACAAAGCTTACCGTAGACCATATATATACACCATAG 54835
OY 1141 ATGTGCTTTTCAAGATTTTCAGAGCGCTGAGAAATCTTTGGAGAGATACATTTGAG 1200
DB 54834 AGCTTTTGTCAAAACCTCCCAAGCGAGAGAAATCTTTGGAGAAATTAACAATCGAG 54775
OY 1201 CAATGGCATTCAGAGTTCAGATTTTGGGACGAGCTGTGAGAGGACACGAGATGCTCC 1260
DB 54774 CCAATGCTTTTAAAGCTACTGTACTGTGATCGACCGGAGAGCAATGAGATGTGTAG 54715
OY 1261 TGAAGCGCTCAGCTGGCTTCTGATCCTCGCTGGAGAGAGAGGCGGCGGCTTTGCCAA 1320
DB 54714 TGAATGAAACGACTGGCTCTGTTACATATATGACGAGGAAAAGAGAGTATACCTCTCCCA 54655

FEATURES
source
1..146081
/organism="Oryza sativa (japonica cultivar-group)"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="1"
/clone="P0469B09"
/join(5902..5980,6229..6272,8134..8295)
/note="hypothetical protein"
/protein_id="BA92397.1"
/db_xref="GI:7228437"
/translation="WGAVERSSFLVYVVFALLITSMAGRKMLINKHVOVSMETS
DDDSMHQGGEDDEMILMWHRIILRVKINDIGTDFPTPAKHAEIEN"
9673..9989
/note="5' LTR"
join(10528..12588,12673..13083,13175..13473,13550..13808,
13901..14083,14241..14306,14743..15045,15216..15479)
/note="EST C28952(C62945) corresponds to a region of the

```

predicted gene.  
Similar to maize transposon MudR mudra protein isolog  
(AC003981) "

/codon\_start=1  
/protein\_id="BAA92398.1"  
/db\_xref="GI:7228438"  
/translation="MTSKGIDRPAERTFQSIYSLNLFRIIDQEVYTVSVVVSAT  
EGYFELMFDSTAMKRVYEMAFERSMPLVIFVSVEKDIINVMQTDVGPINAD  
VGPSPQNEQNEPREQAMGEMADEBERGVIITDEMEREDDNEEDADSDDEGDMV  
ATMANEDSGVLIEDHVPMEYKENEVIDGARVYAHKDEMEKAVHMAVLSDEEFV  
VSTNNVYEVRCMEKEDPMRVAHAKKMDNWKYSIVTEHKCTLGQVKEKRYNTSA  
VASEMYSVVGNIQGEPRKIRHIEENKPYTISYVCAKRAKOKIIMEMGYREASTDN  
LPLRLATIQRNNNTYDHLTFTSVDDTKSLORAFESLGCINAPHCPRVCTD  
LTPMTKRGQILTLIGDGNNOVPMAPAPESRNTSWYPLBEVHIAVVRPNVC  
LIDRRAGKALADYLONGDEKRLPAKPPVRSKMRMGNTYFKKAKHMEIF  
KRLCANQKRAKEMKRLDELTTKQDSQRPOVGDPEPPLGALHDPPTMR  
SGSINFTQMIENEPTKWSLLFTDSSRGVITNNLAVNMMRGVRLPLVIAIV  
EFLHGTQAVFRDVKKIGPSADNNIVFGEQVMRLPHLGPANTLPPYSTVGSXAM  
LLOFTADLHPDADNNSVRSLEAYLLMFGVMFTSTGHADVRLVHARSIVDQ  
PODVQMSKSAVALATYALCEACTKDAGAIJAGCPMLQDMAERFAIGRPVDS  
APYVGRSAKQMEDGPTMGTCRGRGRVAVVGRGYPDVFEEPRLOCPNDIYMPY  
TEBAVAARAPLGLSLCTRDQAVMLTILPMVFDIENPHCPORVWROGLRQVPRNV  
QPTVLPAHSLITRRGGLAGLAPVQCYVDVTLATEYINELPHTBEENTDYLRM  
YLPRTRATETPDAPRPHYAVTDHIFTHRDQYFGADARADISADITAVQRLNR  
GLHLDVGEQRTFEDQEKRAVMKVFSCRAVDIVPAGPDLVCLRAPLASQCLQ  
HRRTTYPRMCNFTIEWMDENQNDGTAYRSETSDYLARKDHECRITAAALEM  
VNPGLPTMRREPRCGRCHVIRSVHRTRGRCFVCPNIVDDEPFIHAMWEYR  
RVEIASLDMRNNPLGLPKWSEPRKCGCRDQWVATSEQTFRRRYFVCPVNDDEV  
CLTIYKIRYDLAKITQTE"  
16155..16469  
/note="3' LTR"  
join(23418..23748,25558..25589,25715..26480,26838..26965)  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAA92399.1"  
/db\_xref="GI:7228439"  
/translation="MGANCCIAAKERPOPCVTEVSAFRNVRHSPMSFRMDNRTHI  
EDIMEMPALFENHSGSIRPETKSGSIAPTSNGSGSPDMFNKLCKHSDKRS  
KIARSDLRGRTSSNSPEAKLSKSLDPTVSAVSDKMSISVSPETAIRADPSRS  
RGHSLPTDASMRKARSPQYQYLOVSDSKTSLRLMEGASPEGRSSMLSYCSN  
DLDAVSHGSGSDQMGMRPTSEMVASQSRMNSVDELLGYSVSKTRSNASMLPTH  
SPDOVCCTCKILKERSWMOELAVVALGCHYHADCDLSITLADAKYPPCV  
CTHGEKCTVYKGLKESKTKKIKPNVIVDNLDSGSKHQKEKIPLEEELYSQARSM  
RELKSSRIGANDMTWLLNDLRTEGFS"  
31026..31191)  
/note="ESTs C96615 (C10106), C26336 (C12127), D21959 (C10106)  
correspond to a region of the predicted gene.  
Similar to NAM (AL021889) "  
/codon\_start=1  
/protein\_id="BAA92400.1"  
/db\_xref="GI:7228440"  
/translation="MESLMDVLPFGFGFHKDTELISHYKKIKHQKLEVEIIPV  
DIYKHIEPDLPAKCDVPTQDNKMFFAARDKYPNGSRSNRAITAGVWKTGDAIK  
MGKQITGKTKLVFHEGRPTGRTEIMIEEYIDRECOAPCKMAVYLCRIKRN  
DWIPNGNELDNDSDPHEPYDAPSEVISTEOLPAAPVGEAPVGAAPEDGTVS  
ATANIPSPSDINDLDMTELFPDFPDQPSIASDLSPDQCVSNVAGALAKV  
OYSPSPNAPVDDTEYLPEDEVYLIHPTGDDPNMLQNPLOQYITQYADVSGLOKE  
ELMSPNAPPSQNSAANNGIIRIRYSKTPETSPVQFKGTAKRKGINKMTSS  
SESINQITKENGSLRVEHQKQADVASLTKSDAKSTELSSNRGLGIRNAPAG  
CSDARMMILVAFALGAVVALHIGQRLGLSGRDQHT"  
38998  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAA92401.1"  
/db\_xref="GI:7228441"  
/translation="MMGHVYLRADTAPDMMANVQPEYRVVDRHGTPEVYVLCRPD  
DRGCRASLMAAMKMLBLYFEGCCYDDPEHGPSTSI"  
complement(39315..39726)  
/note="3' LTR"  
complement(join(40209..40396,40917..43545))  
/note="EST C28952 (C62945) corresponds to a region of the  
predicted gene.

Similar to maize transposon MudR mudra protein isolog.  
(AC003981) "

/codon\_start=1  
/protein\_id="BAA92402.1"  
/db\_xref="GI:7228442"  
/translation="MSSKVTFOIYHGBGNIRFGPDVLDSPVMTSKGIDRPAERTFO  
SIYSWLRGRIQDEYTVMSVVVSATGTYEWEIMPDSTDMARVYEMAFERSMPL  
VIFVSVEKDIINVMQTDVGPINADVGPSPQNEQNEPREQAMGEMADEBERGVI  
ITDEMEREDDNEEDADSDDEGDMVATMANEDSGVLIEDHVPMEYKENEVIDGAR  
VYAHKDEMEKAVHMAVLSDEEFV VSTNNVYEVRCMEKEDPMRVAHAKKMDNWKYS  
IVTEHKCTLGQVKEKRYNTSA VASEMYSVVGNIQGEPRKIRHIEENKPYTISYV  
CAKRAKOKIIMEMGYREASTDN LPLRLATIQRNNNTYDHLTFTSVDDTKSLORAF  
ESLGCINAPHCPRVCTD LTPMTKRGQILTLIGDGNNOVPMAPAPESRNTSWYPL  
BEVHIAVVRPNVC LIDRRAGKALADYLONGDEKRLPAKPPVRSKMRMGNTYFKK  
AKHMEIF KRLCANQKRAKEMKRLDELTTKQDSQRPOVGDPEPPLGALHDPPTMR  
SGSINFTQMIENEPTKWSLLFTDSSRGVITNNLAVNMMRGVRLPLVIAIV  
EFLHGTQAVFRDVKKIGPSADNNIVFGEQVMRLPHLGPANTLPPYSTVGSXAM  
LLOFTADLHPDADNNSVRSLEAYLLMFGVMFTSTGHADVRLVHARSIVDQ  
PODVQMSKSAVALATYALCEACTKDAGAIJAGCPMLQDMAERFAIGRPVDS  
APYVGRSAKQMEDGPTMGTCRGRGRVAVVGRGYPDVFEEPRLOCPNDIYMPY  
TEBAVAARAPLGLSLCTRDQAVMLTILPMVFDIENPHCPORVWROGLRQVPRNV  
QPTVLPAHSLITRRGGLAGLAPVQCYVDVTLATEYINELPHTBEENTDYLRM  
YLPRTRATETPDAPRPHYAVTDHIFTHRDQYFGADARADISADITAVQRLNR  
GLHLDVGEQRTFEDQEKRAVMKVFSCRAVDIVPAGPDLVCLRAPLASQCLQ  
HRRTTYPRMCNFTIEWMDENQNDGTAYRSETSDYLARKDHECRITAAALEM  
VNPGLPTMRREPRCGRCHVIRSVHRTRGRCFVCPNIVDDEPFIHAMWEYR  
RVEIASLDMRNNPLGLPKWSEPRKCGCRDQWVATSEQTFRRRYFVCPVNDDEV  
CLTIYKIRYDLAKITQTE"  
43999..44385)  
complement(43999..44385)  
/note="5' LTR"  
join(44572..44723,44846..44938,46366..46378)  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAA92403.1"  
/db\_xref="GI:7228443"  
/translation="MALPRGVSRPKPMMAAASASAPQPRGHMLVLTGTFVGRVYQ  
RLAQGRVSGTCTSPAKTGLEMLQMDASIDPATSRCW"  
complement(join(48119..48257,48429..48703,49975..50139))  
/note="hypothetical protein"  
/codon\_start=1  
/protein\_id="BAA92404.1"  
/db\_xref="GI:7228444"  
/translation="MALVNDKSDSDVNNPACARAAFLVIGAPKGFVQLARLIGEC  
EPRFRRAKKAITSTCALSLFNSNRTITVLCRKNNRPLIYKESKRKCCRIT  
ASLRSRGDLGCRARLASALDENVPRKILKTSDFLWMLKRCIVINPVHSRTH  
GSSWYSLTDPVAVYRSTRLAGGVIDCTIGQ"  
complement(join(51918..52637,52684..53346))  
/note="Similar to Arabidopsis thaliana chromosome 1 YAC  
YUPB12R sequence, C2-HC type zinc finger protein  
C.e-WY1. (AC002986) "  
/codon\_start=1  
/protein\_id="BAA92405.1"  
/db\_xref="GI:7228445"  
/translation="WASGNGELHVRVVRRLVYKASDESIRPHVLANSNLDLIPRTI  
QVSWFCIYKPSGTGDFHDVVAAPAGLPSLNFHFPFACRIVSNPCSGLEPIHCNQ  
GAEVLVGEADVALSLDYGASVGLILPYAGGVALSVOVVSFAGGFTVAGTNGH  
VVVDGSLMLVSAWSEIASGTLAAGAPNHDSVFRPSPSYASLDEAFPLDG  
ARQADILARQARATRVQVAVLWKRALAAVGSNDACRMVWVWVQRRRLTSS  
PELRAMRSYGVNTVTFPAVAEMTEIQRCLAVYASAMADAIARAAYGEHFOELVW  
VEHKKGRVYITDASVGLSPAVSTAPASBEVDTPFGHMAALPTSSSSRLCT  
GFQYIARSGDGSWIASLMLPRLAALBESDLRLIRKPVTAIVYGLRPAAGSSA  
KRGMITSXI"  
complement(57570..58009)  
/note="3' LTR"  
complement(join(58018..58891,58980..62098))  
/note="Similar to Arabidopsis thaliana chromosome II BAC  
F9013 genomic sequence, putative retroelement pol  
polyprotein. (AC006248) "  
/codon\_start=1  
/protein\_id="BAA92406.1"  
/db\_xref="GI:7228446"  
/translation="WRVNLQAQGLWTADPGYAFREDMALASAILQAVPREMLRGIA  
KHDTAKAADAIKTMRVIVRVEAKEOGFRSESRFERETBEERFAMLTIVADV  
IDMGSGVMEDEHVNKLRLVPRPKYPAVIALBOLDLVKTMLELVEVLSYVSGD  
DESGSGGLYLTBEQWQARVQROBGSGSKGKAGPCGNLRGDRGSGPKGXA  
ATGANSRDIRKVCFCNDEFHVARQCKRKRRRRREANLVQAAEERTILMAHYVG

Query Match 13.2%; Score 224.8; DB 8; Length 146081;  
Best Local Similarity 70.2%; Pred. No. 8e-38;







ATGCTTTGGAAGGATAACAATTGAAGCAATGGCATTCAGTTGCCAGTATTGGGC 122

/note="Similar to gb|Y12575 histone H2A.F/Z from ESTs  
Arabidopsis thaliana. (This gene is cut off.)  
gb|T14101, gb|T42178, gb|T43143, gb|Z26465, gb|Z28692 and

```

/db_xref="taxon:3702"
/chromosome="1"
/clone="F6D8"
/complement<1..400)
/feature="F6D8.1"
/complement[join<1..20,293..400)]
/feature="F6D8.1"
/note="Similar to gb|112576, gene H2A.F/2 from Arabidopsis thaliana. This gene is cut off." ESM
gb|114101, gb|112178, gb|173213, gb|226465, gb|228692 and

```

```

gb|143615 come from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD5592.1"
/db_xref="GI:5903033"
/translation="MSGAGALIMKPSGSKDKKKKPIITRSSRAGLQPLCEIINS"
/gene="F6D8.2"
complement(831..2943)
CDS
/gene="F6D8.2"
complement(join(831..1072,1299..1438,1534..1671,
1782..1957,2322..2403,2509..2644,2826..2943))
/gene="F6D8.2"
/translation="Similar to gb|A010025 unr-interacting protein from
Homo sapiens and contains 3 Pf|00400 WD40 domains. EST.
gb|145021 come from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD5591.1"
/db_xref="GI:5903032"
/translation="MEKKVTIPQVCHGRSPVVDLPYSPITPDGFLISAKDSHPM
LRNGETDMNIGTEBHGKAVMSGLDNNALRAASADPSAKIMDLATGVDLHSEBK
HIVACAFSDTKSLTNGPEKTIIPVDMRLDAPPEVDKSGSRITITMLHSDQTI
LSCTDIGVRLMDYRSKTIQTIETKSVTAESQDGRITTAQSTVFWPANH
GLYKSDPMONTESASLEPESGKFAVGEDWVRVDFYTGEEIGCKNGHGHVHCV
RFTPTGASTASGEDGTIRLWQTPANPENETSSRRVKSVDVEKSLIEGPHINKG
KTAKEPSDT"
complement(449..4802)
CDS
/gene="F6D8.3"
complement(449..4802)
/gene="F6D8.3"
/codon_start=1
/evidence=not_experimental
/product="F6D8.3"
/protein_id="AAD5593.1"
/db_xref="GI:5903034"
/translation="WALITCSALPTIRASGSGSLNPQNRKKSAAWAPLPGLPED
PDYINTESSCTVNPDKTIDISGQKFRRCGTBEKAKQLRKTKLEASFPIDVWYHSA
IASRLASDITGRVED"
<5794..7183
/gene="F6D8.4"
join(<5794..5844,6737..6803,6864..7006,7136..7183)
/gene="F6D8.4"
/translation="Similar to gb|D85381 cytochrome c oxidase subunit
vb precursor from Oryza sativa. ESTs gb|R30504 and
gb|AA598195 come from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD5594.1"
/db_xref="GI:5903035"
/translation="GRRLDIDHPSPFGTSPASVAVQSYFKRNIGCGGEGAGARC
IHLPIICVLIISPETABDGHVFWLMDKGSFPCVCSQYFEHGPDPGHDDDDHHH"
"
/gene="F6D8.5"
join(12444..12474,12682..12793,13387..13423,13523..13592,
13689..13793,13908..14042,14351..14448)
/gene="F6D8.5"
/codon_start=1
/evidence=not_experimental
/product="F6D8.5"
/protein_id="AAD5595.1"
/db_xref="GI:5903036"
/translation="MSYSHQSMGSGRNARAGFERGVYVPRPKHQNTLWMLHGIGD
NGSSSSQLMDSLHLPNKKICTAPSRPVTSIGFTCTAMPDVEGISPDCHDRLKID
ASASHIHLISSEPADIGGFSMGAAISLVATCVCVALGKYGGAHYPLINTQAVYGLSG
WLPADVVYPRFEKKSQSLGMAGRFLMFRPIBG"
15655..17201
/gene="F6D8.6"
join(15655..15664,15795..15849,16134..16170,16281..16350,
16491..16595,16702..16825,17117..17201)
/gene="F6D8.6"
/translation="Similar to F6D8.5."
/codon_start=1
/evidence=not_experimental
"
/gene="F6D8.7"
join(19326..19335,19466..19583,19895..19931,20042..20111,
20228..20332,20406..20531,20797..20924)
/gene="F6D8.7"
/translation="Similar to F6D8.5."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD55622.1"
/db_xref="GI:5903063"
/translation="MASSSRNOSGRKVGERTYISIPPTGVHKATVWLHVDGFTGHS
VPALBSLEPNTKIVPTAPMRPVTSIGEVTTAWCDMPSENNLDIFENLVNPSF
ITSLSFSPDVMKVGIGLGAQALYTSYAFQWVISPQIVIRINGWLPADDDVI
PSAFYKCADSLRMAGFPTLFRQCGSKORLLYFLV1"
22422..24321
/gene="F6D8.8"
join(22422..22431,22558..22669,23113..23149,23253..23322,
23452..23556,23698..23830,24195..24321)
/gene="F6D8.8"
/translation="Similar to F6D8.5."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD55623.1"
/db_xref="GI:5903064"
/translation="MASGSINVSGLERGOINTVYPTGIHKATVWLHVDGNTGFNSLE
PQNLRPNIKVICPTAPRRVTSIGETINMADIAKVSENNODDINYNVEVIT
SLESPQNVKGVAGLGLGAQALYTSYACGAWP2INPQVINGWLPGRKADS
LMAAFPTLPRQCGSKRHLHSFPFPDYFLNTH"
complement(25221..25968)
/gene="F6D8.9"
complement(join(25221..25673,25912..25968))
/gene="F6D8.9"
/translation="Identical to gb|X91919 LEA76 homologue type1 from
Arabidopsis thaliana. ESTs gb|N97082, gb|Z27056 and
gb|Z29902 come from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD5596.1"
/db_xref="GI:5903037"
/translation="MAHQEOSYVAGETRGKQEKTEGAMGTMDKTOAKDQETA
QSAQCKAHEPTAOGAKDTSOAQOTTERAOGESDKGSMSETGEALKNKADAEVT
KETAEAGKERTSGILQGTGEQVKOMAGATDAVKHTLGLRTDEGNKEHVSAPSTTTT
TTRRETORK"
26849..27391
/gene="F6D8.10"
join(26849..26905,27022..27168,27251..27391)
/gene="F6D8.10"
/codon_start=1
/evidence=not_experimental
/product="F6D8.10"
/protein_id="AAD5597.1"
/db_xref="GI:5903038"
/translation="MSSSQSLSHSAGRYGVOLKKEPYLNNVSHAMNONADPHHTSO
SOLSHEDNDNBSLISQASSVIQGTGGQYKNAQADAVKNTLGMSPATNPSFSG
TTRNKGKSKIT"
28814..30691
/gene="F6D8.11"
join(28814..28895,29040..29116,29376..29435,29835..30071,
30159..30221,30327..30395,30472..30517,30621..30691)
/gene="F6D8.11"
/translation="Is a member of the Pf|00364 Biotin-requiring
enzymes family. ESTs gb|F19971 and gb|F19970 come from
this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD55598.1"
/db_xref="GI:5903039"

```

/translation="MKVSLRTLLRSVKAIOISTVPAETEAIDVYKDSDETSTVYN  
THLMPSSEVALISEITLDDSSIAEPELKGFRLYVARKLTDESSPPQIQVVA

Query Match 5.2%; Score 87.8; DB 8; Length 117737;  
Best Local Similarity 53.3%; Pred. No. 3.2e-08;  
Matches 185; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 1110 TTGGGAGTGGGCCCCCTTACTTGGCAGCAATTGATGCTTGTTCAGAAATTCACGGCCGT 1169  
DB 115127 TTGACATTTCGATTTTGTTCGTTATTATTTTGAACCTTGATTTTACAGGAGTT 115068  
QY 1170 GGAGAACTCTTGGAAAGATATACATTGAAGCAATGGCATTTCAAGTCCAGATTTGGCC 1229  
DB 115067 GGATGAACATTGGTGTGAGTGAACATTCAGACCAATGCTTATGACCTTGAGTGGGGA 115008  
QY 1230 ACGGCTGCTGAGAGGACCAACGAGATCCCTCTGAGCGGCTCGACTGCGCTTGTGCATCT 1289  
DB 115007 ACAGATGACAGAGGAAACCAAGAGATGGTTCAACACACATGACTGCTACTTCACTCA 114948  
QY 1290 GCTGGGAGAGAGGGCGCTGCGCTTTCGCAAAAGACATGCTGACACTCGCAGACCGCC 1349  
DB 114947 ATGGGAGATCCGTTAAACAAAGATTACCTATATCTTGTATCTTACCTTACGAAATCCA 114888  
QY 1350 GAGCAGAGGCTCTTCATTCAGGAGGAAAGGCTATGCGAGGATGAAGAAATGTTCAATGAG 1409  
DB 114887 GACGAGAGGCTACGATTGGGAGTGAAGACGCAAAATGTTGAGAAATGTACATGAG 114828  
QY 1410 CACACATGCTGAGAGGATCCGCGGCTGTGAGAGATGTCCTGAG 1456  
DB 114827 CACCATGTATCAAGAGCATTCGTCATATCTCTGCTCAATGCAATGAG 114781

RESULT 8  
AC023010/c 147292 bp DNA linear HTG 18-APR-2000  
LOCUS Homo sapiens clone RP11-17A14, LOW-PASS SEQUENCE SAMPLING.  
DEFINITION AC023010  
ACCESSION AC023010  
VERSION AC023010.2 GI:7582616  
KEYWORDS HTG; HTGS\_PHASB0.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
TITLES Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
JOURNAL 1 (bases 1 to 147292)  
REFERENCE 2 (bases 1 to 147292)

REFERENCE  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
TITLES Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
JOURNAL 1 (bases 1 to 147292)  
REFERENCE 2 (bases 1 to 147292)

REFERENCE  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
TITLES Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
JOURNAL 1 (bases 1 to 147292)  
REFERENCE 2 (bases 1 to 147292)

REFERENCE  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
TITLES Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
JOURNAL 1 (bases 1 to 147292)  
REFERENCE 2 (bases 1 to 147292)

REFERENCE  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,  
TITLES Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
JOURNAL 1 (bases 1 to 147292)  
REFERENCE 2 (bases 1 to 147292)

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIRB  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
-----Project Information  
Center project name: L6614  
Center Clone name: 17\_A\_14

NOTE: This record contains 163 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1 743: contig of 743 bp in length  
\* 744 843: gap of 100 bp  
\* 844 1646: contig of 803 bp in length  
\* 1647 1746: gap of 100 bp  
\* 1747 2505: contig of 759 bp in length  
\* 2506 2605: gap of 100 bp  
\* 2606 3370: contig of 765 bp in length  
\* 3371 3470: gap of 100 bp  
\* 3471 4280: contig of 810 bp in length  
\* 4281 4380: gap of 100 bp  
\* 4381 5149: contig of 769 bp in length  
\* 5150 5249: gap of 100 bp  
\* 5250 5964: contig of 715 bp in length  
\* 5965 6064: gap of 100 bp  
\* 6065 6832: contig of 768 bp in length  
\* 6833 6932: gap of 100 bp  
\* 6933 7656: contig of 724 bp in length  
\* 7657 7756: gap of 100 bp  
\* 7757 8512: contig of 756 bp in length  
\* 8513 8612: gap of 100 bp  
\* 8613 9477: contig of 805 bp in length  
\* 9478 9517: gap of 100 bp  
\* 9518 10329: contig of 812 bp in length  
\* 10330 10429: gap of 100 bp  
\* 10430 11212: contig of 783 bp in length  
\* 11213 11312: gap of 100 bp  
\* 11313 12081: contig of 769 bp in length  
\* 12082 12181: gap of 100 bp  
\* 12182 12894: contig of 713 bp in length  
\* 12895 12994: gap of 100 bp  
\* 12995 13738: contig of 744 bp in length  
\* 13739 13838: gap of 100 bp  
\* 13839 14539: contig of 701 bp in length  
\* 14540 14639: gap of 100 bp  
\* 14640 15395: contig of 756 bp in length  
\* 15396 15495: gap of 100 bp  
\* 15496 16245: contig of 750 bp in length  
\* 16246 16345: gap of 100 bp  
\* 16346 17086: contig of 741 bp in length  
\* 17087 17186: gap of 100 bp  
\* 17187 17926: contig of 740 bp in length  
\* 17927 18026: gap of 100 bp  
\* 18027 18818: contig of 792 bp in length  
\* 18819 18918: gap of 100 bp  
\* 18919 19695: contig of 777 bp in length  
\* 19696 19795: gap of 100 bp  
\* 19796 20570: contig of 775 bp in length  
\* 20571 20670: gap of 100 bp  
\* 20671 21447: contig of 777 bp in length  
\* 21448 21547: gap of 100 bp  
\* 21548 22302: contig of 755 bp in length  
\* 22303 22402: gap of 100 bp  
\* 22403 23163: contig of 761 bp in length  
\* 23164 23263: gap of 100 bp

23264 24028: contig of 765 bp in length  
 \* 24029 24128: gap of 100 bp  
 \* 24129 24872: contig of 744 bp in length  
 \* 24873 24972: gap of 100 bp  
 \* 24973 25721: contig of 749 bp in length  
 \* 25722 25821: gap of 100 bp  
 \* 25822 26599: contig of 778 bp in length  
 \* 26600 26699: gap of 100 bp  
 \* 26700 27460: contig of 761 bp in length  
 \* 27461 27560: gap of 100 bp  
 \* 27561 28341: contig of 781 bp in length  
 \* 28342 28441: gap of 100 bp  
 \* 28442 29199: contig of 758 bp in length  
 \* 29200 29299: gap of 100 bp  
 \* 29300 30083: contig of 784 bp in length  
 \* 30084 30183: gap of 100 bp  
 \* 30184 30928: contig of 745 bp in length  
 \* 30929 31028: gap of 100 bp  
 \* 31029 31786: contig of 758 bp in length  
 \* 31787 31886: gap of 100 bp  
 \* 31887 32640: contig of 754 bp in length  
 \* 32641 32740: gap of 100 bp  
 \* 32741 33484: contig of 744 bp in length  
 \* 33485 33584: gap of 100 bp  
 \* 33585 34353: contig of 769 bp in length  
 \* 34354 34453: gap of 100 bp  
 \* 34454 35198: contig of 745 bp in length  
 \* 35199 35298: gap of 100 bp  
 \* 35299 36061: contig of 763 bp in length  
 \* 36062 36161: gap of 100 bp  
 \* 36162 36924: contig of 763 bp in length  
 \* 36925 37024: gap of 100 bp  
 \* 37025 37800: contig of 776 bp in length  
 \* 37801 37900: gap of 100 bp  
 \* 37901 38663: contig of 763 bp in length  
 \* 38664 38763: gap of 100 bp  
 \* 38764 39531: contig of 768 bp in length  
 \* 39532 39631: gap of 100 bp  
 \* 39632 40387: contig of 756 bp in length  
 \* 40388 40487: gap of 100 bp  
 \* 40488 41273: contig of 786 bp in length  
 \* 41274 41373: gap of 100 bp  
 \* 41374 42145: contig of 772 bp in length  
 \* 42146 42245: gap of 100 bp  
 \* 42246 43022: contig of 777 bp in length  
 \* 43023 43122: gap of 100 bp  
 \* 43123 43864: contig of 742 bp in length  
 \* 43865 43964: gap of 100 bp  
 \* 43965 44694: contig of 730 bp in length  
 \* 44695 44794: gap of 100 bp  
 \* 44795 45464: contig of 670 bp in length  
 \* 45465 45564: gap of 100 bp  
 \* 45565 46349: contig of 785 bp in length  
 \* 46350 46449: gap of 100 bp  
 \* 46450 47230: contig of 781 bp in length  
 \* 47231 47330: gap of 100 bp  
 \* 47331 48102: contig of 777 bp in length  
 \* 48103 48202: gap of 100 bp  
 \* 48203 48981: contig of 779 bp in length  
 \* 48982 49081: gap of 100 bp  
 \* 49082 49867: contig of 786 bp in length  
 \* 49868 49967: gap of 100 bp  
 \* 49968 50741: contig of 774 bp in length  
 \* 50742 50841: gap of 100 bp  
 \* 50842 51608: contig of 767 bp in length  
 \* 51609 51708: gap of 100 bp  
 \* 51709 52475: contig of 767 bp in length  
 \* 52476 52575: gap of 100 bp  
 \* 52576 53113: contig of 738 bp in length  
 \* 53114 53413: gap of 100 bp  
 \* 53414 54164: contig of 751 bp in length  
 \* 54165 54264: gap of 100 bp  
 \* 54265 55003: contig of 739 bp in length

Query Match 3.9%; Score 65.8; DB 2; Length 147292;  
 Best Local Similarity 50.8%; Pred. No. 0.0018;  
 Matches 157; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 6 GCTCGATTGATCCAGCCCAAGTCTCTCGGCGCCAGCGCGTCCAGCCGCGATG 65  
 DB 14464 GCCCGCGGCGGAGGCGCCCGCAGCGCCCGCGCGGAGCGGAGCCACCCAGCGG 14405  
 QY 66 GCAAGACCCCGTGTCTTCGCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 125  
 DB 14404 GAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 14345  
 QY 126 CGAGCCCAAGTCT 185  
 DB 14344 GGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 14285  
 QY 186 TTCT 245  
 DB 14284 GGGCT 14225  
 QY 246 AAGACCGCGCGTCT 305  
 DB 14224 CGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 14165  
 QY 306 GTCTCCCAT 314  
 DB 14164 GCCTGTAAT 14156

RESULT 9  
 AC109537/c 104486 bp DNA linear HTG 13-JUL-2002  
 LOCUS Rattus norvegicus clone CH230-230E2, \*\*\* SEQUENCING IN PROGRESS  
 DEFINITION \*\*\* 60 unordered pieces.  
 ACCESSION AC109537  
 VERSION AC109537.3 GI:21737934  
 KEYWORDS HTG: HTGS PHASE1.  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 104486)

# REFERENCE

Muzny D.M., Adams C., Adio-Oduola B., Ali-ouman F.R., Allen C.,  
 Alshrooks S.L., Amaralunga H.C., Are J.R., Ayale M., Banks T.,  
 Barbara J., Benton J., Blimie K., Blankenburg K., Bonini D.,  
 Bouh J., Bowie S., Brivela M., Brown E., Brown M., Bryant N.P.,  
 Bunney C., Butch P., Burkett C., Butrell K.L., Byrd N.C.,  
 Cartron T.F., Carter M., Cavazos S.R., Chacko J., Chavez D.,  
 Chen G., Chen R., Chen Z., Chowdhry I., Christopoulos C.,  
 Cleveland C.D., Cox C., Coyle M.D., Dathorne S.R., David R.,  
 Davila M.L., Davis C., Davy-Carroll L., Decker D.A.,  
 Delaney K.R., Delgado O., Denn A.L., Ding Y., Dinh H.H.,  
 Douthwaite K.J., Draper H., Dugan-Rocha S., Durbin K.T.,  
 Earhart C., Edgar D., Edwards C.C., Elhaj C., Escotto M.,  
 Falls T., Ferriguto D., Flagg N., Ford P., Foster P., Frantz P.,  
 Gabisti A., Gao J., Garcia A., Garner T., Garza N., Gill R.,

Correll, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlik, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,  
Horn, F., Howard, S., Huber, J., Hult, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S.,  
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
Kratovic, J., Kureishi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,  
Li, J., Li, Z., Lichtenberg, O., Lieu, C., Liu, J., Liu, W., Louised, H.,  
Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzger, M.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbar, K., Morgan, M., Morris, S.,  
Mosier, M., Neale, D., Newton, J., Newton, S., Ogun, M., Okun, G.,  
Nguyen, N., Nickerson, E., Nwokem, S., Ogun, M., Okun, G.,  
Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Saverly, G.,  
Scherrer, S., Scott, G., Shen, H., Shoshitari, N., Sisom, I.,  
Sodergren, E., Sonaike, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Swatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Taneey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczek, R., Woodson, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.

Unpublished  
Direct Submission  
2 (bases 1 to 104486)  
Worley, K.C.

Submitted (05-FEB-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 104486)  
Worley, K.C.

Direct Submission  
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 12, 2002 this sequence version replaced gi:18846856.

----- Genome Center -----  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information -----  
Center project name: GPNJ  
Center clone name: CH230-230E2

----- Summary Statistics -----  
Sequencing vector: Plasmid  
Chemistry: Dye-terminator Big Dye 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 51833 bases at least Q40  
Consensus quality: 55340 bases at least Q30  
Consensus quality: 58186 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_dirft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_dirft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 60 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1020: contig of 1020 bp in length  
\* 1021 1120: gap of unknown length  
\* 1121 2126: contig of 1006 bp in length  
\* 2127 2226: gap of unknown length  
\* 2227 3387: contig of 1161 bp in length  
\* 3388 3487: gap of unknown length

3488 5002: contig of 1515 bp in length  
\* 5003 5102: gap of unknown length  
\* 5103 6894: contig of 1792 bp in length  
\* 6895 6994: gap of unknown length  
\* 6995 8390: contig of 1396 bp in length  
\* 8391 8490: gap of unknown length  
\* 8491 9759: contig of 1269 bp in length  
\* 9760 9859: gap of unknown length  
\* 9860 10937: contig of 1078 bp in length  
\* 10938 11037: gap of unknown length  
\* 11039 12928: contig of 1891 bp in length  
\* 12929 14172: gap of unknown length  
\* 14173 14272: contig of 1144 bp in length  
\* 14273 15297: gap of unknown length  
\* 15298 15397: gap of unknown length  
\* 15398 16905: contig of 1508 bp in length  
\* 16906 17005: gap of unknown length  
\* 17006 18205: contig of 1200 bp in length  
\* 18206 18305: gap of unknown length  
\* 18306 19622: contig of 1317 bp in length  
\* 19623 19722: gap of unknown length  
\* 19723 20976: contig of 1254 bp in length  
\* 20977 21076: gap of unknown length  
\* 21077 22580: contig of 1504 bp in length  
\* 22581 22680: gap of unknown length  
\* 22681 24233: contig of 1553 bp in length  
\* 24234 24333: gap of unknown length  
\* 24334 25471: contig of 1138 bp in length  
\* 25472 25571: gap of unknown length  
\* 25572 26603: contig of 1032 bp in length  
\* 26604 26703: gap of unknown length  
\* 26704 28205: contig of 1502 bp in length  
\* 28206 28305: gap of unknown length  
\* 28306 30496: contig of 2191 bp in length  
\* 30497 30596: gap of unknown length  
\* 30597 32028: contig of 1432 bp in length  
\* 32029 32128: gap of unknown length  
\* 32129 34350: contig of 2222 bp in length  
\* 34351 34450: gap of unknown length  
\* 34451 36149: contig of 1699 bp in length  
\* 36150 36249: gap of unknown length  
\* 36250 37523: contig of 1274 bp in length  
\* 37524 37623: gap of unknown length  
\* 37624 38879: contig of 1256 bp in length  
\* 38880 38979: gap of unknown length  
\* 38980 40254: contig of 1275 bp in length  
\* 40255 40354: gap of unknown length  
\* 40355 41575: contig of 1221 bp in length  
\* 41576 41675: gap of unknown length  
\* 41676 42909: contig of 1234 bp in length  
\* 42910 43009: gap of unknown length  
\* 43010 44310: contig of 1301 bp in length  
\* 44311 44410: gap of unknown length  
\* 44412 45543: contig of 1133 bp in length  
\* 45544 45643: gap of unknown length  
\* 45644 46934: contig of 1291 bp in length  
\* 46935 47034: gap of unknown length  
\* 47035 48142: contig of 1108 bp in length  
\* 48143 48242: gap of unknown length  
\* 48243 49718: contig of 1476 bp in length  
\* 49719 49818: gap of unknown length  
\* 49819 51226: contig of 1408 bp in length  
\* 51227 51326: gap of unknown length  
\* 51327 52490: contig of 1164 bp in length  
\* 52491 52590: gap of unknown length  
\* 52591 53823: contig of 1233 bp in length  
\* 53824 53923: gap of unknown length  
\* 53924 55605: contig of 1682 bp in length  
\* 55606 55705: gap of unknown length  
\* 55706 57465: contig of 1760 bp in length  
\* 57466 57565: gap of unknown length  
\* 57566 58844: contig of 1279 bp in length

*	58845	58944:	gap of unknown length
*	58945	61093:	contig of 2149 bp in length
*	61094	61193:	gap of unknown length
*	61194	62706:	contig of 1513 bp in length
*	62707	62806:	gap of unknown length
*	62807	64024:	contig of 1218 bp in length
*	64025	64124:	gap of unknown length
*	64125	65734:	contig of 1610 bp in length
*	65735	65834:	gap of unknown length
*	65835	67180:	contig of 1346 bp in length
*	67181	67280:	gap of unknown length
*	67281	69652:	contig of 2372 bp in length
*	69653	69752:	gap of unknown length
*	69753	71873:	contig of 2121 bp in length
*	71874	71973:	gap of unknown length
*	71974	73923:	contig of 1950 bp in length
*	73924	74023:	gap of unknown length
*	74024	76036:	contig of 2013 bp in length
*	76037	76136:	gap of unknown length
*	76137	77977:	contig of 1841 bp in length
*	77978	78077:	gap of unknown length
*	78078	80200:	contig of 2123 bp in length
*	80201	80300:	gap of unknown length
*	80301	82529:	contig of 2229 bp in length
*	82530	82629:	gap of unknown length
*	82630	84884:	contig of 2255 bp in length
*	84885	84984:	gap of unknown length
*	84985	86957:	contig of 1973 bp in length

Query Match 3.8% Score 64.4; DB 2; Length 104486;  
Best Local Similarity 53.6%; Pred. No. 0.0036;  
Matches 155; Conservative 0; Mismatches 132; Indels 2; Gaps 1;

Qy	23	GGCCACAGTCTCTCTCGAGCCGACCGCGCGATGACCAAGACCCCGTCCT	82
Db	12630	GGCGACCTCGCTCCACAGCGCCGCGCCTCCGAGCGGCAATCGCGCGCGCG	12571
Qy	83	CGCGGTGCGCGCGTCTCGCGGCGCGCGCGGTTCAACACCGGACCTCTCTCT	142
Db	12570	GGCGAGCGCGCG--CGCGGCGCGCGAGCGCGCGCGCATCCCCCTCTCTCTC	12513
Qy	143	CT	202
Db	12512	CTCTGCT	12453
Qy	203	CTTGCAGACCTTTCAGACGCGCGTGTGAGACCCCGCGCTCTCTCTCTCTCTCT	262
Db	12452	GGCGGCG	12393
Qy	263	CGGAGTCCCT	311
Db	12392	CGGCGGCG	12344

RESULT 10  
NCBI2N19 42273 bp DNA linear PLN 18-JAN-2002  
LOCUS Neurospora crassa DNA linkage group II BAC clone B12N19.  
DEFINITION A1659987  
ACCESSION A1659987  
VERSION 1  
KEYWORDS A1659987.1 GI:18376315  
SOURCE Neurospora crassa.  
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Sordariales; Sordariaceae; Neurospora.  
REFERENCE 1  
AUTHORS Schulte U., Aign V., Hohnselt J., Brandt P., Fartmann B.,  
Holland R., Nyakatura G., Mewes H.W. and Mannhaupt G.  
JOURNAL Unpublished  
REFERENCES 2 (bases 1 to 42273)  
AUTHORS German Neurospora genome project.  
TITLE Direct Submission  
JOURNAL Submitted (18-JAN-2002) MIPS, Institut fuer Bioinformatik,

GSP-Forschungszentrum fuer Umwelt und Gesundheit, GmbH,  
Ingolstaedter Landstrasse 1, D-85764 Neubrandenburg, FRG, E-mail:  
G.Mannhaupt@gsp.de Project Coordinator: Ulrich Schulte, Institute  
of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf,  
E-mail: ulrich-schulte@uni-duesseldorf.de  
BAC clone 12N19 (strain OR74A) is available at the Fungal Genetic  
Stock Center, <http://www.fgsc.net>  
Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,  
<http://www.mwgdna.com>  
Information on performance of analysis and a more detailed  
annotation of this entry and other sequences of linkage groups II  
and V can be viewed at: <http://mips.gsf.de/proj/neurospora>.

FEATURES

source  
1..42273  
/organism="Neurospora crassa"  
/db\_xref="taxon:5141"  
/chromosome="6"

misc\_feature

1..7212  
/note="overlap to BAC B116, EMBL:AL442043, please refer  
to this entry for analysis and annotation"

gene

9830..11707  
/gene="B12N19.020"  
join(9830..10364,10432..11357,11450..11707)  
/gene="B12N19.020"

CDS

9830..11707  
/note="differs in three residues when compared to  
PIR:T47219  
Contains amino acid permease signature  
(CGPASLLIDFAIGMMPNYYALGELAMRP)  
contains EST gb:BF739557, BF739423"  
/codon\_start=1  
/product="amino acid permease NPA1"  
/protein\_id="CAD21063.1"  
/db\_xref="GI:18376316"

translaction="MSDIEKSGASHDKETRYDAATLPAYAGSVAAQOEDPMTRNG  
LNLSFPRPHFGDTEIERTMSRHLNMGISIGAGFGNGSKALHNGSPASLL  
IDPAIIGMNNVYALGELAMRPVTCGPTVYRFRIDPKMGANNVYVQAIVL  
PLELTVCGPTVGNNDITWITTFVVAITIIINVTGTADEEFNMAFRTGATVIG  
LITCLVAVLGGFKRSRDTHTGALMTSPGAFNAGFRGSCVTAFAFSGELVIG  
LAAEAKNPVKSIPGAIKQVWRITTLVILGLFPVGLIIPYDENLIGANPIFNVAS  
PFVIVGKXANTLGFDSFMNVVIVSVLSGVCSGSGSRTLTLAQQGYAPITFYVD  
RSNPLPSVLLIARGLAYVNLAAAGPTVPMILSLSLAALFTGSGICLSHIFRK  
AMALKGISVEIIPFRAAGVWGLGLVIVLVAQFYVAICPGGSGNTVGEFVQ  
YIALPVVIVWIGFIMKRTSMLRVDQIDVDGRREHDMWSINAYQELAAHPANKRN  
FHKLFV"

exon

9830..10364  
/gene="B12N19.020"  
/number=1  
10365..10431  
/gene="B12N19.020"

intron

10432..11357  
/gene="B12N19.020"  
/number=1  
11358..11449  
/gene="B12N19.020"

exon

11358..11449  
/gene="B12N19.020"  
/number=2  
11358..11389  
/note="23 bp poly(c) tandem repeat"

intron

11389..11450  
/gene="B12N19.020"  
/number=3  
11450..14030  
/note="36 bp atatatacatatatacatat tandem repeat"

repeat\_region

14175..14909  
/gene="B12N19.030"  
14175..14909  
/gene="B12N19.030"

gene

14175..14909  
/gene="B12N19.030"  
/note="contains EST gb:BG279737, BG279736"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAD21064.1"  
/db\_xref="GI:18376317"

repeat\_region

14175..14909  
/gene="B12N19.030"  
/note="contains EST gb:BG279737, BG279736"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAD21064.1"  
/db\_xref="GI:18376317"

exon

14175..14909  
/gene="B12N19.030"  
/note="contains EST gb:BG279737, BG279736"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAD21064.1"  
/db\_xref="GI:18376317"

intron

14175..14909  
/gene="B12N19.030"  
/note="contains EST gb:BG279737, BG279736"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAD21064.1"  
/db\_xref="GI:18376317"

repeat\_region

14175..14909  
/gene="B12N19.030"  
/note="contains EST gb:BG279737, BG279736"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAD21064.1"  
/db\_xref="GI:18376317"

gene

14175..14909  
/gene="B12N19.030"  
/note="contains EST gb:BG279737, BG279736"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAD21064.1"  
/db\_xref="GI:18376317"

CDS

14175..14909  
/gene="B12N19.030"  
/note="contains EST gb:BG279737, BG279736"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAD21064.1"  
/db\_xref="GI:18376317"

translation="WGIFDLRPRONGIPRKHTMPPDILSYONDALDKRHSLSM





REFERENCE Sordariales; Sordariaceae; Neurospora.  
AUTHORS 1 Schulte, U., Aign, V., Hohelsel, J., Brandt, P., Fartmann, B.,  
Holland, R., Nyakatura, G., Mewes, H.W. and Mannhaupt, G.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 122151)  
AUTHORS German Neurospora genome project.  
TITLE Direct Submission  
JOURNAL Submitted (18-JAN-2002) MIPS, Institut fuer Bioinformatik,  
GSR-Forschungszentrum fuer Umwelt und Gesundheit, GmbH,  
Ingolstaedter Landstrasse 1, D-85764 Neuberg, FRG, E-mail:  
G.Mannhaupt@fzj.de Project Coordinator: Ulrich Schulte, Institute  
of Biochemistry, Heinrich-Heine-University, D-40225 Duesseldorf,  
E-mail: ulrich-schulte@uni-duesseldorf.de  
COMMENT this contig is an assembly of BAC 23H20 from 1 to: 52608, and BAC  
22H2 from 52609 to: 122151 (strain OR74A); BAC clones are available  
at the Fungal Genetic Stock Center <http://www.fgsc.net>  
<http://www.mwgdb.com>  
Information on performance of analysis and a more detailed  
annotation of this entry and other sequences of linkage groups II  
and V can be viewed at: <http://mips.gsf.de/proj/neurospora>.  
FEATURES  
source Location/Qualifiers  
1. 122151  
/organism="Neurospora crassa"  
/db\_xref="taxon:5141"  
/chromosome="6"  
1. 18481  
/note="overlap to BAC 12N19, please refer to this entry  
for analysis and annotation"  
misc\_feature 21232..22965  
/gene="B23H20.010"  
join(21232..21352,21416..21461,21521..21639,21698..22965)  
/gene="B23H20.010"  
/note="similarity to probable aminocranferase PA4722,  
Pseudomonas aeruginosa, PIR:D83057  
contains EST gb:AA897847, BG278951, BG279936, AA899026,  
BG279935"  
/codon\_start=1  
/product="conserved hypothetical protein"  
/protein\_id="CAD21073.1"  
/db\_xref="GI:18376327"  
/translation="MASHQHPGPGAFHNKRFSSINRTGILPPRRRFSIOEPSATEASS  
KIHRRFRANHGHLPHAGLDATRSSTGVINCSQAGEHGFEEPEKMANIGQAPVD  
DEIEGCFPRPOLDIIVNSREYGPAGIKPIRELVANLYNEMHGRGKESKYTENVAI  
VPGRAGLIRIAAVANNAVYGFIPDYAVVEMSLFKNPAIIPVPSDDVHIHE  
KIAEELAGTSVITISNPPTGRVAVNPRLAIGKRVANILGPKEFIKAGSGSTLD  
CGNTVPRPBAIIPMLDPTLYKAEKALQRIHCDKDFVGRLEIGSILKLVDSIFY  
LWNLWLPGRFISDGLNPFQALCEKVIWVGIFPDINPSRRRLDPSPCHHFRVLSY  
GPTLDILARGCDGIERVVRKFALEAGAVGVPKAPSPRTGEKEPHTVPSAGGER  
EKDEELHL"  
exon 21232..21352  
/gene="B23H20.010"  
/number=1  
intron 21353..21415  
/gene="B23H20.010"  
/number=1  
exon 21416..21461  
/gene="B23H20.010"  
/number=2  
intron 21462..21520  
/gene="B23H20.010"  
/number=2  
exon 21521..21639  
/gene="B23H20.010"  
/number=3  
intron 21640..21697  
/gene="B23H20.010"  
/number=3  
exon 21698..22965  
/gene="B23H20.010"  
/number=4

gene 24117..25698  
/gene="B23H20.020"  
join(24117..24204,24852..24890,25048..25062,25097..25177,  
25263..25297,25447..25518,25585..25698)  
/gene="B23H20.020"  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="CAD21074.1"  
/db\_xref="GI:18376328"  
/translation="MLVSPHGILLPSSQASIQGCGDARKRKPSPREAAEBDCAGS  
KGSATWYFSANDTRESSEHAGRAAGSPFDGDDQKKGKALKDQTALASAGLA  
ASLQDNEALAEAVPDASPTPGAGTTRKRKRVADSGVSSVDSI"  
24117..24204  
/gene="B23H20.020"  
/number=1  
intron 24205..24851  
/gene="B23H20.020"  
/number=1  
exon 24852..24890  
/gene="B23H20.020"  
/number=2  
intron 24891..25047  
/gene="B23H20.020"  
/number=2  
exon 25048..25062  
/gene="B23H20.020"  
/number=3  
intron 25063..25096  
/gene="B23H20.020"  
/number=3  
exon 25097..25177  
/gene="B23H20.020"  
/number=4  
intron 25178..25262  
/gene="B23H20.020"  
/number=4  
exon 25263..25297  
/gene="B23H20.020"  
/number=5  
intron 25298..25446  
/gene="B23H20.020"  
/number=5  
exon 25447..25518  
/gene="B23H20.020"  
/number=6  
intron 25519..25584  
/gene="B23H20.020"  
/number=6  
exon 25585..25698  
/gene="B23H20.020"  
/number=7  
gene 27953..30094  
/gene="B23H20.030"  
complement(join(27953..29206,29267..30094))  
/gene="B23H20.030"  
complement(join(27953..29206,29267..30094))  
/gene="B23H20.030"  
/note="similarity to ferroxidase precursor, YMR058W,  
Saccharomyces cerevisiae, PIR:A55428  
contains multicopper oxidase signatures  
[GWLFFHCHEIEMWHTSGLAVTF]; Multicopper\_Oxidase2  
[HCHIEHMIISGL]"  
/codon\_start=1  
/product="related to cell surface ferroxidase"  
/protein\_id="CAD21075.1"  
/db\_xref="GI:18376329"  
/translation="MRIISRIALLSLTTAAAHAAVTVDVNNVSWITANPDGLAERPYI  
GINNQWPPRIDVQVGRLLIVNLHNSIGDETSJHFGILPMNGSNMDDPVAVVQCPI  
PGASPTVNFIVDQGTWYWSHNSKGYDPGRGLVHDPKAPFEDDELTVISPM  
YHDMSLQPRFMSKVNPTGABVPKSAIMNDTONLIMPVVPRTVLPRIANIGAPAG  
OYIWEVDHRTIVEMDGVYVAPRETMYVAAORCSFLLTAKETSRPFIASMPQ  
TLPDTIPGLINTNVTSLILNPTAPLPDPTPLDSILTELDITLLPQDNLITLPSDDQ  
ITLNVMTNNLASGANYAFENDITVPPRPVTLTYLVSSPNSSSSQMDMTLTPSIYGT

exon	complement (27953..29206) /gene="B23H20.030" /number=1 complement (29207..29266) /gene="B23H20.030" /number=1 complement (29267..30094) /gene="B23H20.030" /number=2 30243 /note="24 bp atgcagtcagtc tandem repeat"
repeat_region	31765..33093 /gene="B23H20.040" join(31765..32576,32637..33093)
CDS	/note="similarity to high-affinity iron permease catf2, candida albicans, TREMBL:AF195776.1 contains EST gb:AA901925, BF072546" /codon_start=1 /product="related to high-affinity iron permease" /protein_id="CAD21076.1" /db_xref="GI:18376330" /translation="WQLFSVPIPLVPFETIETSIYSVLVFLKQTLASPESSSPA EAIASHAATAAADDELKHPSSTTNNNSDSPQQPVAVDLRVNLRRLQILITV LGELIIAIGAVGVFTLVGNWKESGAEILAMEGAFCILASVIIISVGALILLRTP MEKKRGKMIKAMEGEKNKKSKSGRYAGGEDMERGETVERGRGMVVEFKAMERY VMEILPVTVRREVVAIVFAVSFAAPATSIIPPAIVGVISGLVGVLTFFEGST KLOLFESTSLILVALVAGLFSRALAESQOMKAIGSPAVELGGPGSYDIDKESVM HDUSCPVONGGGMGITFNALFGMTNASITVGSVIGDIWLVIIVITSFLMRYREYVNGK WPLPKTEYVDGSEKKSLEGRLSMRGRA"
exon	31765..32576 /gene="B23H20.040" /number=1 32577..32636 /gene="B23H20.040"
intron	/number=1 32637..33093 /gene="B23H20.040"
exon	/number=2 33444..35567
gene	

Query Match 3.4%; Score 58.2; DB 8; Length 122151;  
Best local Similarity 50.9%; Pred. No. 0.079;  
Matches 138; Conservative 0; Mismatches 133; Indels 0; Gaps 0;

Qy	1168 GTGAGAATCTTTGGAAGCATTAACATTGAAGCAATGGCATTCAATTGCCATATTGG	1227
Db	13728 GTAACGAACCTTGATTCGTCCCGCTTAACCAATGTAAAGGCGCGTGCCTGTGG	1378
Qy	1228 GCACGGCTGTCGAGGGAACAACGAGATTCGTCTGAGCGCTGCACATGCGCTTTCGATC	1287
Db	13788 CGCGGATATATGTGTGACCAGCAACGAGACGCTGTGTGAAGGGAAGACGGGTGTTGAGAG	1384
Qy	1288 CTGCTGGGAAGAGGCGCTGCGCTCTTCCAAAAGAACATGTCAGACTCGCAACCCACAG	1347
Db	13848 ACCCGAATGACGTAGGAGGAGTGGCTTAGGATTGATTAAGGTTTTGAATGAGATGGGCGG	1390
Qy	1348 CCGAGCAGAGGATCTTCATGTGGGAAAAAGGGCTATGCGAGGGGTGAAGAAATGTTCATGG	1407
Db	13908 AGGAAGAGTTGAAGAGATGGGAAAAAGAGGGGTGTGAAGGAGGTCAAGGTTAGTTGCGG	1396
Qy	1408 AGCACCATGCTGAGAGATTCGGCGCGGT	1438
Db	13968 ATACCGAATGCGCGAGAGGTTTGAAGAGAT	13998

RESULT 12

AC126306/C	LOCUS	AC126306	152607 bp	DNA	linear	HTG 24-JUL-2002
DEFINITION	Rattus norvegicus clone CH230-207N13, *** SEQUENCING IN PROGRESS					
ACCESSION	AC126306	GI:21699124				
KEYWORDS	HTG; HTGS; PHASE1.					
SOURCE	Rattus norvegicus					
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.					
REFERENCE	1 (bases 1 to 152607)					
AUTHORS	Wuzy,D.M., Adams,C., Adio-Otunola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T., Barbard,J., Benton,J., Bimage,K., Blankenshaw,K., Bonnah,D., Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burich,P., Burnett,C., Burrell,K.L., Byrd,N.C., Cartron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dedereich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,F., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J.J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Koryah,J., Kovar,C., Kratochvil,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., LinchARGE,O., Lien,C., Liu,J., Liu,W., Louisedge,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapa,R., Martin,R., Marinatle,A., Martinez,E., Messay,B., Mihiney,E., McLeod,M., Meador,M., Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokkenwo,S., Ogun,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pichens,R., Plimus,E., Pylun,L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolle,M., Ruiz,S., Savery,G., Scheerer,S., Scott,G., Shen,H., Shooshari,N., Sison,I., Soedergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H., Sulton,A., Swalek,A., Tabor,P., Tanerisa,A., Tamarisita,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Umanli,K., Vasquez,L., Vera,V., Villalongo,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.					
TITLE	Direct Submission					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 152607)					
AUTHORS	Worley,K.C.					
TITLE	Direct Submission					
JOURNAL	Submitted (05-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA					
REFERENCE	3 (bases 1 to 152607)					
AUTHORS	Worley,K.C.					
TITLE	Direct Submission					
JOURNAL	Submitted (24-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA					
COMMENT	----- Genome Center Center: Baylor College of Medicine Center code: BCM Web site: <a href="http://www.hgsc.bcm.tmc.edu/">http://www.hgsc.bcm.tmc.edu/</a> Contact: hgsc-help@bcm.tmc.edu ----- Project Information					

```
Center project name: GZ1Y
Center clone name: CH230-207N13
----- Summary Statistics -----
Sequencing vector: Plasmid:
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 92958 bases at least Q40
Consensus quality: 99800 bases at least Q30
Consensus quality: 103479 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length.
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 65 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1163: contig of 1162 bp in length
1263: gap of unknown length
1283: contig of 1143 bp in length
2406: gap of unknown length
2506: contig of 1264 bp in length
3770: gap of unknown length
3870: contig of 1343 bp in length
5213: gap of unknown length
5313: contig of 1628 bp in length
6941: gap of unknown length
7041: contig of 1065 bp in length
8106: gap of unknown length
8206: contig of 1153 bp in length
9359: gap of unknown length
9459: contig of 1817 bp in length
11276: gap of unknown length
11376: contig of 1388 bp in length
12764: gap of unknown length
12864: contig of 1330 bp in length
14194: gap of unknown length
14294: contig of 1608 bp in length
15902: gap of unknown length
16001: contig of 1652 bp in length
17654: gap of unknown length
17753: contig of 1069 bp in length
18823: gap of unknown length
18923: contig of 1037 bp in length
19860: gap of unknown length
20060: contig of 1615 bp in length
21674: gap of unknown length
21774: contig of 1183 bp in length
22958: gap of unknown length
23057: contig of 1202 bp in length
24259: gap of unknown length
24360: contig of 1615 bp in length
25974: gap of unknown length
26074: contig of 1508 bp in length
27682: gap of unknown length
27882: contig of 1426 bp in length
29109: gap of unknown length
29208: contig of 2167 bp in length
31375: gap of unknown length
31475: contig of 1039 bp in length
32515: gap of unknown length
32614: contig of 1084 bp in length
33698: gap of unknown length
33798: contig of 1516 bp in length
35114: gap of unknown length
35415: contig of 1574 bp in length
36988: gap of unknown length
37088: contig of 1276 bp in length
38365: gap of unknown length
38464: contig of 1795 bp in length
40259: contig of 1795 bp in length
```

```

* 40260: gap of unknown length
* 40360: contig of 1650 bp in length
* 42010: gap of unknown length
* 42110: contig of 1473 bp in length
* 43583: gap of unknown length
* 43683: contig of 1205 bp in length
* 44887: gap of unknown length
* 44988: contig of 1083 bp in length
* 46071: gap of unknown length
* 46171: contig of 1631 bp in length
* 47802: gap of unknown length
* 47902: contig of 1944 bp in length
* 49846: gap of unknown length
* 49946: contig of 1306 bp in length
* 51251: gap of unknown length
* 51352: contig of 2031 bp in length
* 53382: gap of unknown length
* 53483: contig of 2084 bp in length
* 55567: gap of unknown length
* 55667: contig of 2128 bp in length
* 57794: gap of unknown length
* 57894: contig of 1425 bp in length
* 59320: gap of unknown length
* 59420: contig of 2266 bp in length
* 61686: gap of unknown length
* 61786: contig of 2667 bp in length
* 64453: gap of unknown length
* 64553: contig of 1869 bp in length
* 66422: gap of unknown length
* 66522: contig of 2248 bp in length
* 68770: gap of unknown length
* 68869: contig of 2549 bp in length
* 71419: gap of unknown length
* 71519: contig of 1875 bp in length
* 73394: gap of unknown length
* 73494: contig of 1796 bp in length
* 75290: gap of unknown length
* 75390: contig of 3042 bp in length
* 78431: gap of unknown length
* 78532: contig of 2931 bp in length
* 81463: gap of unknown length
* 81563: contig of 2322 bp in length
* 83885: gap of unknown length
* 83985: contig of 3060 bp in length
* 87045: gap of unknown length
* 87145: contig of 2315 bp in length
* 89460: gap of unknown length
* 89560: contig of 2362 bp in length
* 91921: gap of unknown length
* 92021: contig of 2552 bp in length
* 94574: gap of unknown length
* 94674: contig of 3842 bp in length
* 98516: gap of unknown length
* 98616: contig of 3064 bp in length
* 101679: gap of unknown length
* 101779: contig of 3064 bp in length

Query Match
Beet Local Similarity 3.3%; Score 57; DB 2; Length 152607;
Matches 166; Conservative 48.6%; Pred. No. 0.15;
Mismatch 165; Indels 0; Gaps 0;

QY 19 CCGAGCCGACAGCTCTCTCGGCGCCAGCGGCTCGACCGCGCATGACAAAGACCCGCT 78
DB 40025 CCGAGCCGCTCTCTCGGCGCCAGCGGCTCGACCGCGCATGACAAAGACCCGCT 78
QY 79 CCGTTCGCGTGGAGCGGCGTTCGCGGCGCGGCGGCGGTTTCACACCGGACCGAGCTCC 138
DB 39965 CCGTTCGCGGCTTCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39906
QY 139 TCCCTTCCTTCCTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 198
DB 39905 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 39846
QY 199 GTGCCCTGCGAGACCTTCGCGACGCGCGTGGGAGACCCGCGCGCGCTTCACACCGCGCTCG 258
```



CDS

**CDS**

230

DESAKRQSMDEIEREFEAATNSLKQNLDDFSEGDDSAESARRNSMLAIEREFEEA

Matches 117; Conservative 0; Mismatches 101; Indels 0; Gaps

Dib

AF128457

### DEFINITION

VERSION

ORGANISM

[illegible]

## AUTHORS

## REFERENCE

TITLE

## FEATYPES

**sour**

3

mpm

323

[illegible]





mRNA  
complement(<29032..>29790)  
/product="hypothetical protein"  
complement(29032..29790)  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="A027673.1"  
/db\_xref="GI:4680493"  
/translation="MAVDGEMPALESQFNOCAITAEFEAAALPDSTPEPTNVASODE  
ESGAGSESDSDDEVIDIIEGDDGSDADYVTGLPEGMALITTTVDKI  
LALRRRFPMPATGGERLMMWSPEFEELIAEELAAADVAJSDQKPECCAMIAEK  
RHPEGYAAVEETSEMTTICRAFPGYNESSDDDLCDVDANECKCGDDMAN  
EFIEEDDSDSEVDSEEDSEKSCCQRMVRLR"  
complement(30165..30488)  
/product="hypothetical protein"  
complement(join(<31774..32072,32911..>33076))  
/product="hypothetical protein"  
complement(join(31774..32072,32911..33076))  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="A027683.1"  
/db\_xref="GI:4680503"  
/translation="MFOAROPATVYNSEYSDKRTIWMYSMLIGALARSSEF  
PLQSRRLQVAYPRKTSINSCMYTIVGLVHNNFVQLFKSADMAIGDVSW  
LSRQIERLMLIDRPAIRGGEPHRSRVHGVYIVCVVASHHLVLAAL"  
join(<34690..34730,35085..35261,35329..35549,  
38631..>39499)  
/product="hypothetical protein"  
join(34690..34730,35085..35261,35329..35549,38631..39499)  
/codon\_start=1  
/product="hypothetical protein"  
/protein\_id="A027674.1"  
/db\_xref="GI:4680494"  
/translation="WQFVADSLFHLIKQTKSPPTKIAQDTINLVAPKLPFKTSL  
RLROKRTIFGTGLPAPKIRHPHOPASFTLPSPPSLPHSSGTEGGSGHGRRA  
ATRGAVSTEAVRVEVATAGAGGVARLEDPVWREBPSAGOPHSPSLPPLIAT  
HATRIIRPHSSASAGIRPRLSTICIGDRASRIIAPLHCTVFSRSARSRAAA  
AAAGLDLHPDVHARPALRLDGAHPDPQPHGAGAGGDDVDLLGAAGR  
RRRLHVGAAAREEVRLPRDELBEHDAVAVGPRGALGSGAGAGGGG  
CLGRPERHGVRLRVDVREAVVHDAGVGVQDVAGLEVARRRRRRVEVGAS  
HAQPHRDQQLVRYOKAHADQDPRPRIOHPLRLVLT"  
complement(join(<40016..40032,40110..40588,40596..>41539))  
/product="receptor kinase-like protein"  
complement(join(40016..40032,40110..40588,40596..41539))  
/note="similar to Oryza sativa EST clone E2886\_2A found in  
GenBank Accession Number C73087"  
/codon\_start=1  
/product="receptor kinase-like protein"  
/protein\_id="A027675.1"  
/db\_xref="GI:4680495"  
/translation="WTLILAFRLSTFLILANGAIVDPGAAAPPTAAADPTTAA  
ILRLKASLIDPTNALFAMSPSPCDETHRPVQCYNGVLGLRLALNSGDP  
FALSLRLPGHLSINLRNNFSGPLPASIAVRALYLSRNASGFGPDVFAAMSW  
LKLILRLDNDSEGLPAGIAGAPRLOELHDNRLEGVPSKLPATLRLFNVSHNL  
TGVLEAVARFNESAFAAGPGLCGAPSGCAALAPHAHTAMPMSADYFAVQ  
BETSVFVMGIIIMLVILVAGAMVILRODEGTSTASGTPLAGPLRAICAPPAAG  
AASAOIYTMKSGAGAGAGGAGVGAARVYVMSNAAGFFGLPELMKASAE  
VANGCTLSATKAAKNGCTVAVKMDNRVGRALFEHIRMGLRHPNVLSPVGY  
HTRKEKLISEFMRGSLVILHGTGSF"  
complement(join(<42779..42931,43886..44595,45422..45637,  
45851..45988,46130..46266,47466..47502,48912..49126,  
52727..53003,53024..>53135))  
/product="putative sugar transporter protein"  
complement(join(42779..42931,43886..44595,45422..45637,  
45851..45988,46130..46266,47466..47502,48912..49126,  
52727..53003,53024..53135))  
/codon\_start=1  
/product="putative sugar transporter protein"  
/protein\_id="A027676.1"  
/db\_xref="GI:4680496"  
/translation="MAPRIYDVTPACAHGVORLAQTALFAHYKXROLKHKGIN  
DDGEVITINAGVRCRAACVLDGGRRGGGGATVVCPEADAARAGAGAVW  
GAGAAGMLCTAAGTGVPMKRTPEPOLCSPTSTNHOSSANOOLPSPOSHRLIHP  
RALPLPAIARSKLASPRRGTGAPFRRGAGARIIISVEPDSRPNHGRSSNY  
LRAIYIDEMKSTVPSAVASIGYTLGMDFTTVIGNPSALANIHKKEFGIINNGPS

Query Match 3.3%; Score 56; DB 8; Length 77605;  
Best Local Similarity 53.1%; Freq. No. 0.22;  
Matches 119; Conservative 0; Mismatches 105; Indels 0; Gaps 0;  
IDGILAVSVFSGSLAITVPSGSLDMLGRITVSYCYLIIMWTLIPSPMKRIMGALFA  
PSLIVFALIVFYLPSPSPMLVSDGKISLARISLQWLKGDVSDRSAGEIALIADGN  
MILETAIVGHAIVGAVRSQSFLGTSTNOMSRHSTFYWHLDPLVDLIGIHSMSLGA  
GRSYVPVNSFNIVEOBTSREQNDLSLOQREAVASAEKGNNGNDLQASILISQVSA  
ETNDINTSPTESSSYLRHGHSTISGLAQDLISLHDHIDEDDEIHIHALSOPAE  
QY 90 GCGGCGTCCGCGAGAGCCGCGGCGCGGCTTCAACAGCAACCGAGCTCCTCTCTCTC 149  
19574 GTGACGCTCGCGCGCGCGCGCGCGCGCTGCTCCCGCCGCTCAAGCTTCACTTC 19633  
QY 150 CTCG 209  
19634 ATCG 19693  
QY 210 GACCGTTCGAGCG 269  
19694 ATCTTCG 19753  
QY 270 CCCCTCGGTTCTAGAGTCCAGCTCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 313  
Db 19754 CTTTCATGTTCT 19797

Search completed: July 4, 2003, 21:37:10  
Job time : 299 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 14:31:05 ; Search time 283 Seconds  
(without alignments)  
13543.820 Million cell updates/sec

Title: US-09-938-294-2

Perfect score: 1702  
Sequence: 1 atcgcgtcagattgaccc.....gagsgatctcaatcttagc 1702

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- N Geneseq, 101002:\*
- 1: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
  - 2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
  - 3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
  - 4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
  - 5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
  - 6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
  - 7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
  - 8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
  - 9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
  - 10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
  - 11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
  - 12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
  - 13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
  - 14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
  - 15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
  - 16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
  - 17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
  - 18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
  - 19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
  - 20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
  - 21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
  - 22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
  - 23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
  - 24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	589.6	34.6	1715	21	AAc42068
2	154.6	9.1	284	21	AAa31140
3	49.8	2.9	912	21	AAa23717
4	48.4	2.8	1076	20	AAa23007
5	48.4	2.8	1082	24	AAa22912
6	48.2	2.8	856	20	AAa19784
7	48.2	2.8	856	20	AAa20787
8	48.2	2.8	900	21	AAa29385
9	48.2	2.8	912	20	AAa26227

10	48.2	2.8	912	21	AAa288468
11	48.2	2.8	912	22	AAa58896
12	48.2	2.8	912	24	AAa93386
13	48.2	2.8	950	19	AAa17085
14	48.2	2.8	1225	22	AAa21456
15	48.2	2.8	1257	24	AAa18180
16	48.2	2.8	2004	22	AAa09259
17	48.2	2.8	2035	22	AAa09268
18	48.2	2.8	2247	22	AAa09280
19	48.2	2.8	4488	22	AAa09266
20	48.2	2.8	6404	21	AAa09266
21	48.2	2.8	6420	21	AAa09266
22	48.2	2.8	65140	22	AAa09266
23	48.2	2.8	125401	22	AAa09266
24	48.2	2.7	43280	19	AAa19862
25	46.6	2.7	954	19	AAa19862
26	46.6	2.7	1347	19	AAa19864
27	46.6	2.7	1587	19	AAa19863
28	46.6	2.7	2233	19	AAa19861
29	46.6	2.7	5027	19	AAa19860
30	46.4	2.7	21721	20	AAa83427
31	46.4	2.7	23187	21	AAa50273
32	46.4	2.7	23187	22	AAa62331
33	46.2	2.7	4403765	22	AAa199683
34	46.2	2.7	4411529	22	AAa199682
35	46	2.7	1905	18	AAa84863
36	45.6	2.7	2653	21	AAa00338
37	45.6	2.7	44861	24	AAa20000
38	45.2	2.7	1551	22	AAa05562
39	45.2	2.7	1590	24	AAa054274
40	44.6	2.6	27541	22	AAa17185
41	44.6	2.6	4403765	22	AAa199683
42	44.6	2.6	4411529	22	AAa199682
43	44.4	2.6	1216	21	AAa46527
44	44.4	2.6	1440	22	AAa159765
45	44.4	2.6	1440	22	AAa159766

ALIGNMENTS

RESULT 1	
AAc42068	
ID	AAc42068 standard; DNA; 1715 BP.
AC	AAc42068;
XX	
DT	17-OCT-2000 (first entry)
XX	
DE	Arabidopsis thaliana DNA fragment SEQ ID NO: 34167.
XX	
KW	Hybridisation assay; genetic mapping; gene expression control;
KW	protein identification; signal transduction pathway;
KW	metabolic pathway; promoter; termination sequence; ss.
XX	
OS	Arabidopsis thaliana.
XX	
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126284.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.

HSV-1 VP22 peptide  
HSV-1 VP22 coding  
HSV-1 VP22 protein  
Herpes simplex vir  
VP22/57 fusion con  
DNA encoding pcDNA  
VP22-Cre fusion DN  
VP22-CreStreptTag fu  
VP22-F1pe fusion D  
pT7-VPCS vector DN  
Nucleotide sequenc  
Streptomyces nouns  
Tylosane synthase  
Gallus domesticus  
Gallus domesticus  
Gallus domesticus  
Gallus domesticus  
Gallus domesticus  
Human lipolysis st  
Human lipolysis st  
Human leptin fragm  
Mycobacterium tube  
Mycobacterium tube  
Aspergillus niger  
Wheat raffinose sy  
DNA encoding pyrid  
Human secreted pro  
Human ovarian anti  
Streptomyces nouns  
Mycobacterium tube  
Zea mays DNA fragm  
Human polynucleoti

PR	19-APR-1999	99US-0130077
PR	21-APR-1999	99US-0130449
PR	23-APR-1999	99US-0130510
PR	28-APR-1999	99US-0130891
PR	28-APR-1999	99US-0131449
PR	30-APR-1999	99US-0132048
PR	30-APR-1999	99US-0132407
PR	04-MAY-1999	99US-0132471
PR	04-MAY-1999	99US-0132484
PR	06-MAY-1999	99US-0132486
PR	06-MAY-1999	99US-0132487
PR	07-MAY-1999	99US-0132683
PR	11-MAY-1999	99US-0134256
PR	14-MAY-1999	99US-0134218
PR	14-MAY-1999	99US-0134219
PR	14-MAY-1999	99US-0134721
PR	14-MAY-1999	99US-0134770
PR	18-MAY-1999	99US-0134768
PR	19-MAY-1999	99US-0134941
PR	20-MAY-1999	99US-0135124
PR	21-MAY-1999	99US-0135523
PR	24-MAY-1999	99US-0135629
PR	25-MAY-1999	99US-0136021
PR	27-MAY-1999	99US-0136592
PR	28-MAY-1999	99US-0136782
PR	01-JUN-1999	99US-0137222
PR	03-JUN-1999	99US-0137528
PR	04-JUN-1999	99US-0137502
PR	07-JUN-1999	99US-0137724
PR	08-JUN-1999	99US-0138094
PR	10-JUN-1999	99US-0138547
PR	10-JUN-1999	99US-0138647
PR	14-JUN-1999	99US-0139119
PR	16-JUN-1999	99US-0139453
PR	17-JUN-1999	99US-0139492
PR	18-JUN-1999	99US-0139454
PR	18-JUN-1999	99US-0139455
PR	18-JUN-1999	99US-0139456
PR	18-JUN-1999	99US-0139457
PR	18-JUN-1999	99US-0139458
PR	18-JUN-1999	99US-0139459
PR	18-JUN-1999	99US-0139460
PR	18-JUN-1999	99US-0139461
PR	18-JUN-1999	99US-0139462
PR	18-JUN-1999	99US-0139463
PR	18-JUN-1999	99US-0139464
PR	26-JUN-1999	99US-0140091
PR	28-JUN-1999	99US-0141287
PR	01-JUL-1999	99US-0141842
PR	01-JUL-1999	99US-0142154
PR	02-JUL-1999	99US-0142055
PR	06-JUL-1999	99US-0142390
PR	08-JUL-1999	99US-0142803
PR	09-JUL-1999	99US-0142920
PR	12-JUL-1999	99US-0142977
PR	13-JUL-1999	99US-0143542
PR	14-JUL-1999	99US-0143624
PR	15-JUL-1999	99US-0144005
PR	16-JUL-1999	99US-0144086
PR	19-JUL-1999	99US-0144325
PR	19-JUL-1999	99US-0144331
PR	19-JUL-1999	99US-0144332
PR	19-JUL-1999	99US-0144333
PR	19-JUL-1999	99US-0144334

[illegible]

```

PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159367.
PR 14-OCT-1999; 99US-0159368.
PR 14-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161932.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 34.6%; Score 589.6; DB 21; Length 1715;

Best Local Similarity 68.1%; Pred. No. 2,2e-156;

Matches 818; Conservative 2; Mismatches 382; Indels 0; Gaps 0;

```

QY 266 GAGTCCCTCGGATTCATGAGTCCAGGCTGCTCTCTCCCATGAGCTTCCT 325
DB 351 GAATCCTCTAGATTTCATGAAATCGAAGCTGTTTGTATGATCCCGAGCTATCT 410
QY 326 CTCTGGTGGTCCACTTTTATCTGATGGAATGACATTTCTTCTGAGGCAATGTTGGCTCGCA 385
DB 411 CTGAVTSGACCTTTTGTCTATGATGGAATGAGCTTTTGTATGAGGAGTTGGTGTGA 470
QY 386 AGTGTGTGATACAAACAGAGATCAGAGAAACAAAGATGCACTATAGCTTGA 445
DB 471 CGTTGTTTGATCACTATCAAAAGCAGCTTGAAGATGAAAGTTGCTATAGCTTGA 530
QY 446 GCATAGGATGTTGAACATGAGTGCAGGTTTTTACAGCTAGAGACAGAGGCAATTGA 505
DB 531 ACACAGAGATGTTGATGAGAGAGTTGAGGTTATCTCTGCAAGGATCAGAAAGCTGAGA 590
QY 506 TATTGCTCTTAAAGCTGATCTGTTATCTTAAACATGCTGCTGTTGGCAAGTGGCTGA 565
DB 591 TACATCTCTCAAGGCTGATTTGATTTGTTCTTATATCTGCTGCTGAGAAATGGCTGA 650
QY 566 CCTGTTCTGAAGATCATGTTCTTAAAGCTTTCGGAAGATTTGTGTGATCTCATGA 625
DB 651 TGCCTGTTCTTAAAGAAATGTTGTTAAAGTTCTTAAAGTCTTATGTTGATTCATGA 710
QY 626 AATGCTGGCATTTCTTAAAGTTGAATATGTCAAACATCTTCCCTTTGTTGCTGAGC 665
DB 711 GATGAGAGACACTATTTCAATGCTATTTGTCAAACCTCCCTTTGTGTCAGGAGC 770
QY 686 CATGATGATTTCTCATACAGGCTGATTTGAAATGAGAGATGAGCATGCGCTGAA 745
DB 771 CATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 830
QY 746 AATACAGATCCCAAACTTATGTTGTCACCTGGGGAATGTAAGAACTTAAAGT 805
DB 831 GATTAAATGCTCCAAATCTTACGTTGTCACCTGAGAAATGCAAGAAATTAATGAGT 890
QY 806 TGTGAGAGCAATGTCGAGAGAGATCTTACGGGAAACATTTGTTGAAATCCCTTGAAGT 865
DB 891 AGCTGAGATAGTGTGTCAGAGAGATTTCTCGTAGACCTCCGAAATCTCTTGAAGT 950
QY 866 ACGAGTGAAGATCTCTGTTGCAATTAATAACGATGATCAAGAGAAAGGACAGA 925
DB 951 GCGAATGAGAGACTTACTTTGGCATTAATATGTTATGCTGCTGAGAAAGGCGCAGA 1010

```

```

QY 926 CTATTTCTTCAAGATTTTATCAGGCTTTGCACTCATTCACAGAGAACTTAAGT 985
DB 1011 TCTGTTCTCCAGGCTCTCCACGAAAGTCTTGAAAGATTCAGAAAGAACTTCAGT 1070
QY 986 GCTTGAATATCATGCTGATGTTGTTGAGAGTATGTTATGCTCAGACCAATTTGAGAC 1045
DB 1071 ACACAAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1130
QY 1046 TCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1105
DB 1131 AGAGCTACGCAATTTTGTCCGAGAAAGAACTTGAAACTTTGTCCACTGCTGACAA 1190
QY 1106 GACATTTGAGAGGCTGCTTCACTTGGAGCAATGATGCTGTTGATTTCTCAGG 1165
DB 1191 AACTTTAAGCTTACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1250
QY 1166 CCGTGAAGAAATGCTTTGAGAGATTAATTAAGCAATGCAATTTCAAGTTGCAAT 1225
DB 1251 CAGAGAGAAATGCTTTGAGAGATTAATTAAGCAATGCAATTTGCAATGCTGATCT 1310
QY 1226 GGGACGAGCTGCTGAGAGGACCAAGAGATGCTGCTGAGACGCTGCACTGCTGCA 1285
DB 1311 TGTATCTGACGCGGAGAGAAATGAGATTTGATGATGATGATGATGATGATGATGAT 1370
QY 1286 TCTGCTGAGAGAGAGGCTGAGGCTCTTTCGAAAGAAATCTGCACTGCAAGCA 1345
DB 1371 TAGTGCAGGAGAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1430
QY 1346 CCGCAGAGAGAGGCTGATGAGGAGAAAGGCTTATGAGAGGCTGAGAGAAATGTTAT 1405
DB 1431 AGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1490
QY 1406 GGAGCAGCAATGCTGAGAGATGCGGCGGCTTGAAGATGCTCTGAGAGAAATCA 1465
DB 1491 GGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1550
QY 1466 GG 1467
DB 1551 GG 1552

RESULT 2
AAA31140
ID AAA31140 standard; DNA; 284 BP.
XX
AC AAA31140;
XX
DT 05-JUL-2000 (first entry)
XX
DE Plant microsatellite marker #101.
XX
KW Plant microsatellite sequence; core repeat sequence; detection; probe;
XX
KW DNA polymorphism; genome mapping; physical mapping; fingerprinting;
XX
KW variety identification; genetic variability evaluation; primer; ss.
XX
OS Eucalyptus grandis.
XX
XX
XX WO9967421-A1.
XX
XX PD 29-DEC-1999.
XX
XX PF 25-JUN-1999; 99WO-NZ00092.
XX
XX PR 25-JUN-1998; 98US-0105307.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD & FLETCHER.
XX
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Havukala ID, Bloksberg LN, Glenn M.
XX
XX WPI, 2000-116958/10.
XX
XX PT New plant microsatellite markers and associated flanking species for

```

OS	Synthetic.
XX	Herpes simplex virus-1.
XX	MO200022110-A2.
XX	
PD	20-APR-2000.
PF	08-OCT-1999; 99WO-US2
PR	09-OCT-1998; 98US-010
PPA	(HARD) HARVARD COLLEGE
XX	Zhou P, Howley P,
PI	

KM Barley; alpha amylase leader sequence; plant expression;  
KW organophosphorus hydrolase; OPH; environmental detoxification;



QY 304 TCGTCTCCCATGAG 317  
 XX |||||  
 DB 196 TCGGCTCCCGCAG 209

RESULT 6  
 AA219784  
 ID AA219784 standard; DNA; 856 BP.  
 XX

AC AA219784;  
 XX

DT 06-DEC-1999 (first entry)  
 XX

DE Herpes simplex virus type 1 (HSV-1) VP22 DNA.  
 XX

KW Cytochrome; targeting; localisation; cancer; tumour; prodnrg; reduction;  
 XX nucleus; de.  
 OS Herpes simplex virus type 1.  
 OS Synthetic.

XX Key Location/Qualifiers  
 FH CDS 1..856  
 FT /\*tag= a  
 FT /product= "HSV-1 tegument protein VP22"  
 FT /\*tag= b  
 FT 750..751  
 FT /\*tag= c  
 FT /note= "Apparent 50 bp deletion between these bases  
 FT which alters the reading frame"

XX MO9945127-A2.  
 XX

PD 10-SEP-1999.  
 XX

PR 05-MAR-1999; 99WO-GB00674.  
 XX

PR 06-MAR-1998; 98GB-0004841.  
 XX

PR 19-AUG-1998; 98GB-0018103.  
 XX

PR 29-JAN-1999; 99GB-0002081.  
 XX

PA (OXFO-) OXFORD BIOMEDICA UK LTD.  
 XX

PI Stratford IJ, Patterson AV, Kingsman SM, Kan O, Griffiths L;  
 XX Mitochondrion K;  
 XX

DR WPI; 1999-551046/46.  
 XX

DR P-PSDB; AA42292.  
 XX

PT New prodnrg activating agent targeted to selected cells or tissues,  
 XX particularly hypoxic cells, for treating e.g. tumors -

PS Example 7; Fig 3; 187pp; English.  
 XX

CC This sequence represents DNA encoding a Herpes simplex virus type 1  
 CC (HSV-1) VP22 tegument protein, which is involved in transcellular  
 CC localisation. VP22 can be fused to cytochrome P450 reductase (P450R)  
 CC derivatives such as anchorless P450R (AAV42287) or FN fragment  
 CC (AAV42288). This enables the fusion protein to be delivered to  
 CC neighbouring cells where it is then transported to the nucleus. Many  
 CC drugs' sites of action are in the nucleus, rather than the cytoplasm,  
 CC where P450R normally functions. P450R or its derivatives can be used to  
 CC activate prodnrgs to their active form via reduction. Administration of a  
 CC prodnrg is useful where the active drug may be metabolised before it  
 CC reaches its site of action or where the active drug is cytotoxic, e.g.,  
 CC anticancer drugs. Targeted delivery of such prodnrg activators allows a  
 CC reduction in dose of the prodnrg, and thus of systemic side-effects.  
 CC P450R derivative fusion proteins, or vectors that express them, are  
 CC specifically used to treat tumours, inflammation, atherosclerosis and  
 CC muscular dystrophy, but may also be used to treat many other conditions,  
 CC e.g., cerebral malaria, rheumatoid arthritis, or conditions associated  
 CC with hypoxia, ischaemia or hypoglycaemia, or to deliver antibiotics,  
 CC antiviral agents, analgesics, anaesthetics, anti-inflammatory agents,

CC antineoplastic agents and diagnostic agents.  
 XX  
 SQ Sequence 856 BP; 137 A; 339 C; 270 G; 110 T; 0 other;

Query Match 2.8%; Score 48.2; DB 20; Length 856;  
 Best Local Similarity 47.2%; Pred. No. 0.0049;  
 Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 3 TCGGCTCGAGTTGATCCGAGCCCAAGTCTCTCTCGGAGCCCAAGCGTCCGACCGGCG 62  
 DB 322 TCCGAGAGGAGCGGAGACGACACACACACACGCGCCCGGCGCCCGGAGACCGAGCGGTG 381  
 QY 63 ATGGCAAGACCCCGCTTTGCGGTGGCGGCGGTGCGCGAGCGCGCGCGGTTCAC 122  
 DB 382 GCGACTAAGGCGCCCGCGCGCGCGCGCGAGACACCGCGCGAGAAATCGGCGCAG 441  
 QY 123 AACCGACCCGAGCT 182  
 DB 442 CCAGATCCGCGGCACTCCGACGCGCCCGGCTGAGCGGCGGCAACCGATCCAAAGCA 501  
 QY 183 GGGTTCT 242  
 DB 502 CCGCGGAGGAGGTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 561  
 QY 243 CTCAACCGCGCGGTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 302  
 DB 562 CCATGACCGCGCGCGGCGGCGGCTTTAACAAGCGCGCTTCTGCGCGCGGCGGCGGCG 621  
 QY 303 CTGCTCTCC 311  
 DB 622 CTGCGCGCC 630

RESULT 7  
 AA207807  
 ID AA207807 standard; DNA; 856 BP.  
 XX

AC AA207807;  
 XX

DT 23-NOV-1999 (first entry)  
 XX

DE HSV-1 tegument protein VP22 encoding DNA.  
 XX

KW Prodng; localization domain; tumor-selective antibody; cytochrome P450;  
 XX prodnrg activating domain; modified hematopoietic stem cell; MHC; tumor;  
 KW inflammation; atherosclerosis; muscular dystrophy; cerebral malaria;  
 KW rheumatoid arthritis; hypoxia; ischaemia; hypoglycaemia; VP22; HSV;  
 XX tegument protein; ss.

OS Herpes simplex virus type 1.  
 XX

XX Key Location/Qualifiers  
 FH CDS 1..854  
 FT /\*tag= a  
 FT /transl\_except= (pos:751..754; aa:Leu, Leu, Gln,  
 FT Arg, Ala, Asn, Glu, Leu, Val, Asn,  
 FT Pro, Asp, Val, Val, Gln, Asp, Val)  
 FT /note= "a line of DNA sequence is possibly missing;  
 FT the corresponding protein has 17 amino acid  
 FT residues respective to this region for which  
 FT a DNA sequence of only 4 basepairs is indicated"

XX MO9945126-A2.  
 XX

PD 10-SEP-1999.  
 XX

PR 05-MAR-1999; 99WO-GB00672.  
 XX

PR 06-MAR-1998; 98GB-0004841.  
 XX

PR 19-AUG-1998; 98GB-0018103.  
 XX

PR 29-JAN-1999; 99GB-0002081.  
 XX

PA (OXFO-) OXFORD BIOMEDICA UK LTD.  
 XX





QY 63 ATGGCAAAAGACCCGCTGTTGCGGCTGCGCGGTCGCGAGCGCGCGGCTTCAC 122  
 DB 373 GGCACCTAAGAGCCCGCGGCGCGGAGACACCCGCGGAGAAATGCGGCCAG 432  
 QY 123 AACCGGACCCGACCTCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 182  
 DB 433 CCAGAACTCGCGCGACCTCCAGACGCGCGCGCTCGACGCGCGCAACCCGATCCAGAGA 492  
 QY 183 GGGTTCCT 242  
 DB 493 CCGCGCGACGAGGCTGCGCGAGAAAGCTGCACTTTAGACGCGCGCGCAACCCCGACGCG 552  
 QY 243 CTCGAACACCGCGCTGCGCGAGCGAGTCCCTCGGCGTTATAGGCTCCAGCTCTGCTC 302  
 DB 553 CCATGAGACCCCGCGGCTGCGCGCTTTAAAGAGCGCTCTCTCTGCGCGCGGTCGCGCGC 612  
 QY 303 CTCGCTCTCC 311  
 DB 613 CTGGCGGCGC 621

## RESULT 9

AXX26227  
 ID AAX26227 standard; DNA, 912 BP.

AC AAX26227;

DT 25-MAY-1999 (first entry)

DE HIV-1 VP22 polypeptide encoding DNA.

XX Cyclin-dependent kinase; CDK; CDK/cyclin complex; inhibitory; restenosis;  
 KW CDK-binding motif; endothelialisation; fusion protein; therapeutic; acne;  
 KW intracellular; transcellular; transcytosis; vascular wound; repair; hair;  
 KW smooth muscle; cardiovascular; arteriosclerotic; fibrotic disorder;  
 KW cellular proliferation; rheumatoid arthritis; diabetes; cirrhosis; graft;  
 KW tumour; inflammation; neurodegeneration; periodontal; spermatogenesis;  
 KW tachycardia; HIV-1; ss.

OS Human immunodeficiency virus type 1.

PN WO9906540-A2.

PD 11-FEB-1999.

PF 29-JUL-1998; 98WO-US15759.

PR 29-JUL-1997; 97US-0902572.

PA (MITO-) MITOTIX INC.

PI Beach DH, Gyuris J, Lamphere L;

DR WPI; 1999-153770/13.

DR P-PSDB; AAM95099.

XX Fusion and chimeric proteins including cyclin-dependent kinase  
 PT binding motif - used for regulation of cell proliferation and  
 PT differentiation for treatment of, e.g. vascular injury, cancers,  
 PT fibrosis and neurodegeneration

PS Example 2; Page 26; 88pp; English.

XX The invention relates to novel inhibitors of cyclin-dependent kinases  
 CC (CDKs), particularly CDK/cyclin complexes. It provides a recombinant  
 CC transfection system (A) that comprises: (i) first gene construct  
 CC comprising a sequence encoding an inhibitory polypeptide containing at  
 CC least one CDK-binding motif for binding and inhibiting activity of a  
 CC CDK, linked to a transcription regulator functional in eukaryotic cells;  
 CC (ii) second gene construct comprising a sequence encoding a polypeptide  
 CC that promotes endothelialisation, and (iii) a gene delivery composition  
 CC for delivering the GCS to a cell for transfection. Also provided are

CC nucleic acids encoding a fusion protein (FP) containing: (i) a  
 CC therapeutic polypeptide sequence (TP) from an intracellular protein that  
 CC alters a cellular process when FP enters the cell, and (ii) a  
 CC transcellular polypeptide sequence (TCP) that promotes transcytosis of  
 CC FP. The FP consists of at least one CDK-binding motif and a TCP. See  
 CC AAX26220 for detailed uses of the recombinant transfection system. The  
 CC CKI polypeptides are engineered to include any of the peptides shown in  
 CC AAM95097-100 encoded by the DNA sequences AAX26225-228.

XX Sequence 912 BP; 149 A; 354 C; 287 G; 122 T; 0 other;

Query Match 2.8%; Score 48.2; DB 20; Length 912;

Best Local Similarity 47.2%; Fred. No. 0.0051;

Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 3 TCGGCTGAGATTGATCCGAGCCGACAGTCTCTCTCGGCGCCACCGCGTCCGACCGGCG 62

DB 325 TCCGAGAGGCGCGAGCGACACCCACACCGCGCGCGCGCGCGCGCGCGCGCGCG 364

QY 63 ATGGCAAAAGACCCGCTGTTGCGGCTGCGCGGTCGCGAGCGCGCGGCTTCAC 122

DB 385 GGCACCTAAGAGCCCGCGGCGCGGAGACACCCGCGGAGAAATGCGGCCAG 444

QY 123 AACCGGACCCGACCTCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 182

DB 445 CCAGAACTCGCGCGACCTCCAGACGCGCGCGCTCGACGCGCGCAACCCGATCCAGAGA 504

QY 183 GGGTTCCT 242

DB 505 CCGCGCGACGAGGCTGCGCGAGAAAGCTGCACTTTAGACGCGCGCGCAACCCCGACGCG 564

QY 243 CTCGAACACCGCGCTGCGCGAGCGAGTCCCTCGGCGTTATAGGCTCCAGCTCTGCTC 302

DB 555 CCATGAGACCCCGCGGCTGCGCGCTTTAAAGAGCGCTCTCTGCGCGCGGTCGCGCGC 624

QY 303 CTCGCTCTCC 311

DB 625 CTGGCGGCGC 633

## RESULT 10

AAZ88468  
 ID AAZ88468 standard; DNA, 912 BP.

AC AAZ88468;

DT 10-MAY-2000 (first entry)

DE HSV-1 VP22 peptide encoding nucleotide sequence.

XX Papillomavirus; PV; infection; cell proliferation; E2; peptidomimetic;  
 KW E1; antiviral; virucide; cytostatic; antiproliferative; dermatological;  
 KW preneoplastic lesion; neoplastic lesion; cutaneous lesion; wart;  
 KW epidermodysplasia verruciformis; anorectal carcinoma; ss.

OS Herpes simplex virus type 1.

PN WO200001720-A2.

PD 13-JAN-2000.

PF 02-JUL-1999; 99WO-US15144.

PR 02-JUL-1998; 98US-0091661.

PA (HARD ) HARVARD COLLEGE.

PI Howley P, Benson J, Kasukawa H;

DR WPI; 2000-171001/15.

DR P-PSDB; AAY9877.

PT Use of papillomavirus E2 protein peptidomimetics for treating

PT papillomavirus-infected cells and papillomavirus-induced conditions in  
PT mammals by inhibiting E1-E2 interaction  
PS  
XX Disclosure; Page 42; 110pp; English.

CC The present invention describes the use of a small organic compound (A)  
CC which competitively inhibits interaction of a papillomavirus (PV) E2  
CC protein with a PV E1 protein for treating a cell infected with PV or a  
CC mammal with a PV-induced condition. (A) has antiviral, virucide,  
CC cyostatic, antiproliferative and dermatological activities. Methods  
CC from the present invention can be used to treat PV-induced conditions  
CC including growth of PV preneoplastic and neoplastic lesions, cutaneous  
CC lesions chosen from warts and other benign cutaneous lesions, plantar  
CC warts (verruca plantaris), common warts (verruca plana), Butcher's  
CC common warts, flat warts, genital warts (condyloma acuminatum) and  
CC epidermodyplasia verruciformis, laryngeal, oral, pharyngeal,  
CC oesophagial and other upper airway papilloma or vaginal, cervical,  
CC vulvar, penile and anorectal carcinoma. The E2 inhibitors may also be  
CC used to treat epithelial and internal fibropapillomas in animals.  
CC The present sequence represents a nucleotide sequence used in the  
CC exemplification of the present invention.

SQ Sequence 912 BP; 149 A; 354 C; 287 G; 122 T; 0 other;

Query Match 2.8%; Score 48.2; DB 21; Length 912;  
Best Local Similarity 47.2%; Pred. No. 0.0051;  
Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 3 TCGGCGTCAAGTTTATCCAGCCCAAGTCTCTCTCGGGGCGACCGGTCGACCGGGG 62  
DB TCGGAGGGGCGCGACACACACACCCCGGGGCGCGCGCGCGCGCGCGCGCGCGG 384  
QY 63 ATGGCAAAAGACCCGCTGTTGCGGGTGGCGGCGTCCGAGAGCGCGCGCGGTTAC 122  
DB GCGACTAAGGCG 444  
QY 123 AACCGGACCGACT 182  
DB 445 CCAGAAATCCGCGCACTCCCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 504  
QY 183 GGGTTCCT 242  
DB 505 CCGCGCGAGGGGCTGGCGCAAAAGCTGCACTTTAGCACTCCCGCGCGCGCGCG 564  
QY 243 CTCACACCGCGCGTGGCGCAGGAGTCCCTCGGGTTATAGAGTCCAGCTCGGTC 302  
DB 565 CCAATGACCCCGCGGCTGGCGGCTTTAAAGAGCGGTCTTCTGCGCGCGGTCG 624  
QY 303 CTCGTCTCC 311  
DB 625 CTGGCGGCGC 633

RESULT 11

AAFS8996  
ID AAFS8996 standard; DNA; 912 BP.

AAFS8996;

05-NOV-2001 (first entry)

HSV-1 VP22 coding sequence.

KW Co-activator domain; P300/CBP KIX domain; erythrocythemia; skin disease;  
KW polycythemia; haemoglobinopathy; cell differentiation; ulcer; cancer;  
KW neurological condition; neurodegenerative disease; immune disease;  
KW diabetes; ds.

OS Synthetic.

WO200118036-A2.

15-MAR-2001.

XX 31-AUG-2000; 2000MO-US24010.  
XX  
XX 03-SEP-1999; 99US-0152402.  
XX

PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.  
PA (JOSL-) JOSLIN DIABETES CENT INC.

PI Frangioni JV, Cantley LC, Montminy MR;

DR WPI; 2001-273380/28.

DR P-PSDB; AAB60910.

PT Identifying co-activator domain specific transcriptional activators by  
PT contacting a target domain of a selected transcription factor with a  
PT peptide display library, where the identified binding peptides are  
PT useful for reducing hypoglycemia -

PS Disclosure; Page 78; 156pp; English.

CC The present invention describes a method of identifying the co-activator  
CC domain of specific transcriptional activators, involving contacting the target  
CC domain of a selected transcription factor with a peptide display library,  
CC and identifying those sequences which bind to the target domain. In  
CC particular, those which bind to the KIX domain of P300/CBP are useful.  
CC The peptides can be used in the treatment of diseases related to aberrant  
CC KIX-dependent gene transcription, including erythrocythemia,  
CC polycythemia, haemoglobinopathies, to regulate cell differentiation, to  
CC treat neurological diseases, immunological diseases, diabetes, ulcers,  
CC skin diseases and cancer, and to aid wound healing. The present sequence  
CC is a coding sequence described in the exemplification of the invention.

SQ Sequence 912 BP; 149 A; 354 C; 287 G; 122 T; 0 other;

Query Match 2.8%; Score 48.2; DB 22; Length 912;  
Best Local Similarity 47.2%; Pred. No. 0.0051;  
Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 3 TCGGCGTCAAGTTTATCCAGCCCAAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 62  
DB TCGGAGGGGCGCGACACACACACCCCGGGGCGCGCGCGCGCGCGCGCGCGCGG 384  
QY 63 ATGGCAAAAGACCCGCTGTTGCGGGTGGCGGCGTCCGAGAGCGCGCGCGGTTAC 122  
DB 325 TCGGAGGGGCGCGACACACACACCCCGGGGCGCGCGCGCGCGCGCGCGCGG 384  
QY 123 AACCGGACCGACT 182  
DB 385 GCGACTAAGGCG 444  
QY 183 GGGTTCCT 242  
DB 445 CCAGAAATCCGCGCACTCCCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 504  
QY 243 CTCACACCGCGCGTGGCGCAGGAGTCCCTCGGGTTATAGAGTCCAGCTCGGTC 302  
DB 565 CCAATGACCCCGCGGCTGGCGGCTTTAAAGAGCGGTCTTCTGCGCGCGGTCG 624  
QY 303 CTCGTCTCC 311  
DB 625 CTGGCGGCGC 633

RESULT 12

ABA93386  
ID ABA93386 standard; DNA; 912 BP.

ABA93386;

22-APR-2002 (first entry)

HSV-1 VP22 protein encoding DNA.





PT New nucleic acids encoding fusion polypeptide comprising intercalin  
 PT transport polypeptide linked to antigenic polypeptide, useful as  
 PT therapeutic vaccine for cancer and major chronic viral infections -  
 XX  
 PS Disclosure; Fig 7; 102pp; English.  
 YX



QY TCCCGGAGAGCGGGGCGGGTTCACACCGAGCCAGCTCTCTCTCCCTCCCGCTG 156

Db TCCCGCTCTCTCCCGTCTCTGTCTCTCTCCCGGCTCTGTATCTCTCTCTCCCGCTCCCGG 470

QY CCGTGGACGCTCCGATCACAGGAGTTCCTCTCTCCGCGGCGCTGACAGACCTT 216

Db TCTCTGCTCTCTCTCCCGATCTCTCCCGCTCCCGGTCCTGTATCTCTCTCCCGGCTCTG 530

QY GGCAACGCGCGTGGGAGACCCCGCGCCTCAACACGCGCTGTGCACGGGAGATCTCTGT 276

Db TCTCTCTCTCCCGTCTCCCGTCTCTCTCTCTCCCGTCTCCCGTCTCCCGCTCTCCCGTCTC 590

QY GGTTCATGAGGTCCAGACGTCTGTCTCTGCTCTCCCATGAGCTCTCCCTC 326

Db TGTCTCTCTCTCCCGTCTCTGTCTCTCTCCCGTCTCTCCCGTCTCTCTCCCGTCT 640

```

1  RESULT 3
2  US-09-347-504-11
3  ; Sequence 11, Application US/09347504
4  ; Patent No. 6398075
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Howley, Peter M.
7  ; APPLICANT: Benson, John
8  ; APPLICANT: Kasukawa, Hiroaki
9  ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
10 ; TITLE OF INVENTION: PAPILLOMAVIRUS-INFECTED CELLS
11 ; FILE REFERENCE: HMV-041.01
12 ; CURRENT APPLICATION NUMBER: US/09/347,504
13 ; CURRENT FILING DATE: 1999-07-02
14 ; NUMBER OF SEQ ID NOS: 79
15 ; SOFTWARE: PatentIn Ver. 2.1
16 ; SEQ ID NO 11
17 ; LENGTH: 912
18 ; TYPE: DNA
19 ; ORGANISM: Artificial Sequence
20 ; FEATURE:
21 ; OTHER INFORMATION: Description of Artificial Sequence: Ndel-EcoRI
22 ; OTHER INFORMATION: Fragment
23 ; US-09-347-504-11

```

[illegible]

```

RESULT 4
US-09-230-421-1
; Sequence 1, Application US/09230421
; Patent No. 6200577
; GENERAL INFORMATION:
; APPLICANT: Medical Research Council
; TITLE OF INVENTION: ANTI-HERPESVIRAL AGENTS AND ASSAYS
; TITLE OF INVENTION: THEREXOR
; FILE REFERENCE: P18189C
; CURRENT APPLICATION NUMBER: US/09/230,421
; CURRENT FILING DATE: 1999-01-25
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 950
; TYPE: DNA
; ORGANISM: HERPESVIRUS TYPE 1
; US-09-230-421-1

```

Query Match	2.8%;	Score 48.2;	DB 4;	Length 950;
Best Local Similarity	47.2%;	Pred. No. 0.0011;		
Matches 146; Conservative	0;	Mismatches 163;	Indels 0;	Gaps 0

QY	3	TCGAGCTCGAATTGATTCGAGAGCCGACAGTCTCTCTCGAGGCGCCAGACGGGTCCGACACGGGCG	62
Db	366	TCGGAGGGGGGCGGAGACGACACCCACACCGGCCCCCGGGGCCCGGAAACCCAGCGGGT	425
QY	63	ATGCGAAGAAAGACCCCGTCTGTGTGGGGGTGGCGGTGCGCGGAGCGCGGGCCGGTTCCAC	122
Db	426	GCAGACTTAAGGCCCCCGGGCCCCCGGGCGGGAGACACCGCGCGGAGAAATTCGCCCCAG	485
QY	123	AACCGGACCGAGCTCTCTCTCTCTCTCTCTCTGTGGGCGGTGCGACGCTTCGCAATCCAGACA	182
Db	486	CCGAGATTCGCCCGCATCTCCAGAGCGCCCCCGGTGAGCGGCGCCAAACCCGATCCAGACA	545
QY	183	GGGTTCTCTCTCGCGGATGACCTGCGAGACCTTGCGAGCGGCGGTGGGGAGCCCGCGCGCC	242
Db	546	CCCGCGCAGGGGCGTGGCCAGAAAGCTGACCTTAAACACCGCCCCCGCAATCCGACGCG	605
QY	243	CTCAACACCGCCGCTCGCGCAGCGGGAGTCCCTCGGGTTCAATAGAGTTCAGACTGTGTCTC	302
Db	606	CCATGACACCCCGGGGTGGCGGCTTTAAACAGCGCGTCTTTCGCGCGCGGTGGGGCGC	665
QY	303	CTCGTCTTC	311
Db	666	CTGGCGGCGC	674

RESULT 5  
 US-08-232-463-14  
 : Sequence 14, Application US/08232463  
 : Patent No. 5670367  
 GENERAL INFORMATION:  
 APPLICANT: DORNER, F.  
 APPLICANT: SCHEIFLINGER, F.  
 TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
 NUMBER OF SEQUENCES: 52



```

TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII(DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 43280 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 816..14234
FEATURE:
NAME/KEY: CDS
LOCATION: 14351..19945
FEATURE:
NAME/KEY: CDS
LOCATION: 20010..31199
FEATURE:
NAME/KEY: CDS
LOCATION: 31232..36067
FEATURE:
NAME/KEY: CDS
LOCATION: 36249..41774
US-08-804-227C-1

Query Match 2.7%, Score 46.8; DB 2; Length 43280;
Best Local Similarity 48.8%; Pred. No. 0.018;
Matches 126; Conservative 0; Mismatches 132; Indels 0; Gaps 0;

Oy 50 CGTCCGACCGGGGAGTGGCAAAAGACCCCGTCGTCGCGAGTGGCGGCGGCGGAGCGCG 109
Db 18070 CGGCTCCCGGAGACCGGAAACCGCCGGGAGCACACCTGGGCGGTGGCCGGGCGCGG 18129
Oy 110 CGGCGCGGTTCACAAACCGAGACCAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 169
Db 18130 GCACTGTTCACCGGGGAGAACCCCGACACCAACCCGACTCGCTTCGCGGTGCGGACAC 18189
Oy 170 GCGATCCACAGAGAGGTTCTCTCTCGCGAGTCCCTCGGAGACCTTGTGCGAGCGGCGTGG 229
Db 18190 CGGACCGCGCCCTTCGCGCTGATGGGGATGTGCGGGATAGGGACCGGACCGCCCGG 18249
Oy 230 GGACCCCGCGCCCTCAACACACCGCGTGGCCAGCGGAGTCCCTCGGTTTCATGAGATC 289
Db 18250 GGAATCGCTCCCAAGGCGCTGCGCTCGTTCGACGAGTGGAGCGACGCGCGGCGGCGG 18309
Oy 290 CAAGCTGTGCTCTCGT 307
Db 18310 CGAAGCGAGGCTCTGCTT 18327

```

## RESULT 7

US-08-680-506-5  
; Sequence 5, Application US/08680506C  
; Patent No. 6008013  
; GENERAL INFORMATION:  
; APPLICANT: Reynolds, Paul R.  
; TITLE OF INVENTION: CHONDROCYTE PROTEINS  
; FILE REFERENCE: 176/60091  
; CURRENT APPLICATION NUMBER: US/08/680,506C  
; CURRENT FILING DATE: 1996-07-08  
; EARLIER APPLICATION NUMBER: 60/021,672  
; EARLIER FILING DATE: 1996-07-05  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 5  
; LENGTH: 954  
; TYPE: DNA  
; ORGANISM: Gallus gallus  
US-08-680-506-5

## Query Match

2.7%; Score 46.6; DB 3; Length 954;  
Best Local Similarity 53.6%; Pred. No. 0.003;  
Matches 97; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 79 CGTTCGGGTTGGGCGGCGGTGCGCGAGGCGCGGCGGTTTCACACCGGACCGAGCTCC 138  
Db 125 CGCTGTCCCTCGGGCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 184  
Qy 139 TCCT 198  
Db 185 AGCTCCGCGCGCGGTGTCT 244  
Qy 199 GTGCGCTTCGAGACCTTCGCGAGCGCGCGTGGGACCCCGCGCGCGCGCGCGCGCGTGG 258  
Db 245 TGGCCGAGCTGCGACGCGCGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 304  
Qy 259 C 259  
Db 305 C 305

## RESULT 8

US-08-680-506-8  
; Sequence 8, Application US/08680506C  
; Patent No. 6008013  
; GENERAL INFORMATION:  
; APPLICANT: Reynolds, Paul R.  
; TITLE OF INVENTION: CHONDROCYTE PROTEINS  
; FILE REFERENCE: 176/60091  
; CURRENT APPLICATION NUMBER: US/08/680,506C  
; CURRENT FILING DATE: 1996-07-08  
; EARLIER APPLICATION NUMBER: 60/021,672  
; EARLIER FILING DATE: 1996-07-05  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 8  
; LENGTH: 1347  
; TYPE: DNA  
; ORGANISM: Gallus gallus  
US-08-680-506-8

Query Match 2.7%; Score 46.6; DB 3; Length 1347;  
Best Local Similarity 53.6%; Pred. No. 0.0035;  
Matches 97; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 79 CGTTCGGGTTGGGCGGCGGTGCGCGAGGCGCGGCGGTTTCACACCGGACCGAGCTCC 138  
Db 125 CGCTGTCCCTCGGGCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 184  
Qy 139 TCCT 198  
Db 185 AGCTCCGCGCGGTGTCT 244

Qy 199 GTGCGCTTCGAGACCTTCGCGAGCGCGCGGAGCCCGCGCGCGCTTCACACCGGACCGAGCTCC 258

Db 245 TGGCCGAGCTGCGACGCGCGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 304

Qy 259 C 259  
Db 305 C 305

## RESULT 9

US-08-680-506-6  
; Sequence 6, Application US/08680506C  
; Patent No. 6008013  
; GENERAL INFORMATION:  
; APPLICANT: Reynolds, Paul R.  
; TITLE OF INVENTION: CHONDROCYTE PROTEINS  
; FILE REFERENCE: 176/60091  
; CURRENT APPLICATION NUMBER: US/08/680,506C  
; CURRENT FILING DATE: 1996-07-08  
; EARLIER APPLICATION NUMBER: 60/021,672  
; EARLIER FILING DATE: 1996-07-05  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 6  
; LENGTH: 1587  
; TYPE: DNA  
; ORGANISM: Gallus gallus  
US-08-680-506-6

## Query Match

2.7%; Score 46.6; DB 3; Length 1587;  
Best Local Similarity 53.6%; Pred. No. 0.0038;  
Matches 97; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 79 CGTTCGGGTTGGGCGGCGGTGCGCGAGGCGCGGCGGTTTCACACCGGACCGAGCTCC 138  
Db 205 CGCTGTCCCTCGGGCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 264  
Qy 139 TCCT 198  
Db 265 AGCTCCGCGCGGTGTCT 324  
Qy 199 GTGCGCTTCGAGACCTTCGCGAGCGCGCGTGGGACCCCGCGCGCGCGCGCGCGTGG 258  
Db 325 TGGCCGAGCTGCGACGCGCGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 384  
Qy 259 C 259  
Db 385 C 385

## RESULT 10

US-08-680-506-4  
; Sequence 4, Application US/08680506C  
; Patent No. 6008013  
; GENERAL INFORMATION:  
; APPLICANT: Reynolds, Paul R.  
; TITLE OF INVENTION: CHONDROCYTE PROTEINS  
; FILE REFERENCE: 176/60091  
; CURRENT APPLICATION NUMBER: US/08/680,506C  
; CURRENT FILING DATE: 1996-07-08  
; EARLIER APPLICATION NUMBER: 60/021,672  
; EARLIER FILING DATE: 1996-07-05  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 2233  
; TYPE: DNA  
; ORGANISM: Gallus gallus  
US-08-680-506-4

Query Match 2.7%; Score 46.6; DB 3; Length 2233;  
Best Local Similarity 53.6%; Pred. No. 0.0045;  
Matches 97; Conservative 0; Mismatches 84; Indels 0; Gaps 0;





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 21:37:15 ; Search time 211 Seconds  
(without alignments)

12591.504 Million cell updates/sec

Title: US-09-938-294-2

Perfect score: 1702  
Sequence: 1 atccggctcgagttgattcc.....gaggagattcaattctagt 1702

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_NA.\*

1: /cgn2\_6/ptodata/2/pubna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/2/pubna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/2/pubna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/2/pubna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/2/pubna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/2/pubna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/2/pubna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/2/pubna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/2/pubna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/2/pubna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/2/pubna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/2/pubna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/2/pubna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/2/pubna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	262	15.4	276	10 US-09-923-876-6035	Sequence 6035, Ap
2	257	15.1	257	10 US-09-923-876-6078	Sequence 6078, Ap
3	52.4	3.1	2987	9 US-10-128-714-250	Sequence 250, App
4	52.4	3.1	3495	9 US-10-128-714-5250	Sequence 5250, App
5	49.8	2.9	1326	9 US-10-156-761-3022	Sequence 3022, Ap
6	49.8	2.9	9025608	9 US-10-156-761-1	Sequence 1, Appl1
7	48.2	2.8	912	8 US-08-902-572-13	Sequence 13, Appl1
8	48.2	2.8	912	9 US-10-161-499-11	Sequence 11, Appl1
9	48.2	2.8	912	10 US-09-862-179A-40	Sequence 40, Appl1
10	48	2.8	918	9 US-10-156-761-5655	Sequence 5655, Ap
11	48	2.8	9025608	9 US-10-156-761-1	Sequence 1, Appl1
12	47	2.8	520	9 US-10-184-644-332	Sequence 332, App
13	47	2.8	520	9 US-10-184-634-332	Sequence 332, App
14	46.2	2.7	1038	9 US-10-156-761-1894	Sequence 1894, Ap
15	46	2.7	1905	9 US-09-990-385-2	Sequence 2, Appl1
16	45.6	2.7	594	9 US-10-123-155-10	Sequence 10, Appl1
17	45.4	2.7	1431	9 US-10-128-714-7250	Sequence 7250, Ap
18	45.4	2.7	1495	9 US-10-128-714-6250	Sequence 6250, Ap
19	45	2.6	621	9 US-10-156-761-4870	Sequence 4870, Ap

20	45	2.6	978	9 US-10-156-761-1779	Sequence 1779, Ap
21	44.8	2.6	1662	9 US-10-156-761-3483	Sequence 3483, Ap
22	44.6	2.6	738	9 US-10-156-761-276	Sequence 276, App
23	44.4	2.6	2480	9 US-10-098-841-184	Sequence 184, App
24	44.4	2.6	2480	9 US-10-098-841-186	Sequence 186, App
25	44.4	2.6	2531	9 US-10-098-841-185	Sequence 185, App
26	44.2	2.6	2855	9 US-09-904-420A-1	Sequence 1, Appl1
27	43.8	2.6	41936	10 US-09-967-768A-116	Sequence 116, App
28	43.8	2.5	1143	9 US-10-156-761-5707	Sequence 5707, App
29	43	2.5	849	9 US-10-156-761-7075	Sequence 7075, App
30	43	2.5	1140	9 US-10-156-761-6044	Sequence 6044, App
31	42.8	2.5	1440	9 US-10-156-761-6190	Sequence 6190, App
32	42.8	2.5	2334	9 US-10-156-761-2734	Sequence 2734, App
33	42.8	2.5	2790	9 US-10-300-834-5	Sequence 5, Appl1
34	42.8	2.5	3432	9 US-10-300-834-4	Sequence 4, Appl1
35	42.6	2.5	1428	9 US-10-156-761-5428	Sequence 5428, App
36	42.4	2.5	390	10 US-09-960-352-14023	Sequence 14023, A
37	42.4	2.5	395	10 US-09-960-352-6638	Sequence 6638, App
38	42.2	2.5	4257	9 US-09-825-288A-1	Sequence 1, Appl1
39	42	2.5	5025	10 US-09-960-253-176	Sequence 176, App
40	42	2.5	42999	9 US-09-799-462A-17	Sequence 17, App1
41	42	2.5	42999	9 US-10-125-767-17	Sequence 17, App1
42	42	2.5	42999	9 US-09-836-911A-17	Sequence 17, App1
43	42	2.5	42999	9 US-10-151-081-17	Sequence 17, App1
44	42	2.5	42999	9 US-10-287-313-17	Sequence 17, App1
45	42	2.5	42999	9 US-10-219-694-17	Sequence 17, App1

## ALIGNMENTS

RESULT 1  
US-09-923-876-6035  
Sequence 6035, Application US/09923876  
Patent No. US20020013958A1  
GENERAL INFORMATION:  
APPLICANT: Laligudi, Raghunath V.  
APPLICANT: Kamigaki, Laura Y. (Ito)  
APPLICANT: Sherman, Bradley K.  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING  
FILE REFERENCE: PL-0012-1 CON  
CURRENT APPLICATION NUMBER: US/09/923, 876  
CURRENT FILING DATE: 2001-08-06  
PRIOR APPLICATION NUMBER: 09/298,329  
PRIOR FILING DATE: 1999-04-21  
PRIOR APPLICATION NUMBER: 60/085,331  
PRIOR FILING DATE: 1998-05-05  
NUMBER OF SEQ ID NOS: 6332  
SOFTWARE: PERL Program  
SEQ ID NO 6035  
LENGTH: 276  
TYPE: DNA  
ORGANISM: Zea mays  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Inocyte ID No. US20020013958A1 700458234H1  
NAME/KEY: unsure  
LOCATION: 200, 275  
OTHER INFORMATION: a, t, c, g, or other  
US-09-923-876-6035

Query Match 15.4%; Score 262; DB 10; Length 276;  
Best Local Similarity 99.3%; Pred. No. 1.1e-72;  
Matches 273; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY	1034	CAATTGGACCTCGCTTACTGCTGTTGTGTGAAGACACGATTCAGACCGGTGCA	1093
DB	1	CAATTGGACCTCGCTTACTGCTGTTGTGTGAAGACACGATTCAGACCGGTGCA	60
QY	1094	TTTGTGACACAGACATTTGGAGTGGAGCCCTTACTTGGACGACATTTGATGCTTTTCA	1153
DB	61	TTTGTGACACAGACATTTGGAGTGGAGCCCTTACTTGGACGACATTTGATGCTTTTCA	120

Qy	1154	GAATTCACAGGCGCTGAGAAATGCTTTGGAAAGATAACATTTGAAGCAATGGGATTCAA	1213
Db	121	GAATCTCAGGGCCGTGAGAAATGCTTTGGAAAGATAACATTTGAAGCAATGGGATTCAA	180
Qy	1214	GTTCGCAAGATATTTGGGACAGGCTGTGAGAGGAAACAAGAGATGTCTCTGAGACGCTGCAC	1273
Db	181	GTTCGCAAGATATTTGGGACAGGCTGTGAGAGGAAACAAGAGATGTCTCTGAGACGCTGCAC	240
Qy	1274	TGGCTTCTTGCAATCTCTCTGAGAGAGAGGCGGTGG	1308
Db	241	TGGCC--TCTGCAATCTCTCTGAGAGAGAGGCGGTGG	274

```

RESULT 2
US-09-923-876-6078
Sequence 6078, Application US/09923876
Patent No. US20020013958A1
GENERAL INFORMATION:
APPLICANT: Lalgudi, Raghunath V.
APPLICANT: Kamigaki, Laura Y. (lto)
APPLICANT: Sherman, Bradley K.
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
FILE REFERENCE: PL-0012-1 CON
CURRENT APPLICATION NUMBER: US/09/923_876
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: 09/298_329
PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: 60/085_331
PRIOR FILING DATE: 1998-05-05
NUMBER OF SEQ. ID NOS: 6332
SOFTWARE: PERL Program
SEQ ID NO 6078
LENGTH: 257
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. US20020013958A1 700458322H1
US-09-923-876-6078

```

Query Match	15.1%	Score 257;	DB 10;	Length 257;
Best Local Similarity	100.0%	Pred. No. 46-71;		
Matches 257;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1035	AAATTGGAGACTCAGTTAGCTGATCTTTGTGTGAAGAACACAGATTCAAGCCGTTGCAT	1094	
Db	1	AAATTGGAGACTCAGTTAGCTGATCTTTGTGTGAAGAACACAGATTCAATACCGTTGCAT	60	
Qy	1095	TTTGTGACAAAGACATTGGCAGTGGCCCTCTTAATTGGGCAAGATTAGTGTGTTTCAG	1154	
Db	61	TTTGTGACAAAGACATTGGCAGTGGCCCTCTTAATTGGGCAAGATTAGTGTGTTTCAG	120	
Qy	1155	AAATTTCAGGGCCGTGGAGATGCTTTTGGAAAGATTAACATTGAGCAATTGCAAG	1214	
Db	121	AAATTTCAGGGCCGTGGAGATGCTTTTGGAAAGATTAACATTGAGCAATTGCAAG	180	
Qy	1215	TTGCGAGATTGGGCAAGGCTGTGAGAGGAGACACAGGAGATCGCTGGACGGGTCGACT	1274	
Db	181	TTGCGAGATTGGGCAAGGCTGTGAGAGGAGACACAGGAGATCGCTGGACGGGTCGACT	240	
Qy	1275	GGCCTTTCGCATCTGCG	1291	
Db	241	GGCCTTTCGCATCTGCG	257	

```

RESULT 3
US-10-128-714-250
; Sequence 250, Application US/10128714
; Publication No. US20030119013A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Bo
; APPLICANT: Hu, Wengli
; APPLICANT: Tishkoff, Daniel

```

```

? APPLICANT: Zamudio, Carlos
? APPLICANT: Broshkin, Alexey M
? APPLICANT: Lemieux, Sebastien M
? TITLE OF INVENTION: Identification of Essential Genes in Aspergillus fumigatus and
? TITLE OF INVENTION: Methods of Use
? FILE REFERENCE: 10192-018-999
? CURRENT APPLICATION NUMBER: US/10/128, 714
? CURRENT FILING DATE: 2002-04-23
? PRIOR APPLICATION NUMBER: US 60/285,697
? PRIOR FILING DATE: 2001-04-23
? PRIOR APPLICATION NUMBER: US 60/287,066
? PRIOR FILING DATE: 2001-04-27
? PRIOR APPLICATION NUMBER: US 60/295,890
? PRIOR FILING DATE: 2001-06-05
? PRIOR APPLICATION NUMBER: US 60/303,899
? PRIOR FILING DATE: 2001-07-09
? PRIOR APPLICATION NUMBER: US 60/316,362
? PRIOR FILING DATE: 2001-08-31
? NUMBER OF SEQ ID NOS: 8603
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 250
? LENGTH: 2987
? TYPE: DNA
? ORGANISM: Aspergillus fumigatus
US-10-128-714-250

```

Query Match	Similarity	3.1%;	Score 52.4;	DB 9;	Length 2987;
Best Local	Similarity	48.1%;	Pred. 0.3.2e-05;		
Matches	149;	Conservative	0;	Mismatches	161;
				Indels	0;
				Gaps	0;
Qy	1173	GAATGCTTTGGAAGATTAACATTGAAAGCAATGAGCATTCAGATTGGCGACGATTGGGCACG	1233	CGATTTGTCGACGATTGGGACG	
Db	2223	GAACACTTTGGTATATTGTCGCCGCTGAAGCCATGACCGCCGCGAGATCCCGGTTCTGCGATCA	2282		
Qy	1233	GCTGCTGAGAGGACCAACGAGATGCTGCTTGAACGGCTCGACTGSCCTTCTGCACTCGCT	1292		
Db	2283	AATACACGGCGGGCCATTAAAGAACTATCGTGAAGCCGAAACCGGGCTGGCTCGCGAGCGCA	2342		
Qy	1293	GGGAGAGAGGGGCGTGGCGGCTCTTGCAAAAGAACATGTCAGAACTGSCAAAGCCACAGCCGAG	1352		
Db	2343	ACGGTTTGTCCGGATTGGACTGTCGGTATGAGACCGGGTTCTTCAACAGATGAGACCAAG	2402		
Qy	1353	CAGAGGGTCTCCATGGGGGGGAAAAAGGCGTATGCGACAGGTGAAGAAATGTTCAATGAGCAC	1412		
Db	2403	GACTTGGACCCGATGTCACACCGGTAGACAGAGGAGAGAGATTCCTCGTCACT	2462		
Qy	1413	CACATGCGCTGAGAGATTCGCGGCGGCTGTTGAAGATGTGCTGGAGAAATCAACAGAGCAC	1472		
Db	2463	GCAATGGGTGACAGGCTCCAGAGAGAGAAATCACCGATATGCTCAGCAAGAAAGACGGCCG	2522		
Qy	1473	TCCAGGTCTT	1482		
Db	2523	TTCCAGGGTT	2532		

RESULT 4  
 US-10-128-714-5250  
 : Sequence 5250, Application US/10128714  
 : Publication No. US20030119013A1  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: Uiang, Bo  
 : APPLICANT: Hu, Wengqi  
 : APPLICANT: Tishkoff, Daniel  
 : APPLICANT: Zamudio, Carlos  
 : APPLICANT: Eroshtkin, Alexey M  
 : APPLICANT: Lemieux, Sebastien M  
 :  
 : TITLE OF INVENTION: Identification of Essential Genes in *Aspergillus fumigatus* and  
 : FILE REFERENCE: 10182-018-999  
 :  
 : CURRENT APPLICATION NUMBER: US/10/128, 714  
 :  
 : CURRENT FILING DATE: 2002-04-23  
 :  
 : PRIOR APPLICATION NUMBER: US 60/285,697  
 :  
 : PRIOR FILING DATE: 2001-04-23

```

; PRIOR APPLICATION NUMBER: US 60/287,066
; PRIOR FILING DATE: 2001-04-27
; PRIOR APPLICATION NUMBER: US 60/295,890
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: US 60/303,899
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/316,362
; PRIOR FILING DATE: 2001-08-31
; NUMBER OF SEQ ID NOS: 8603
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5250
; LENGTH: 3495
; TYPE: DNA
; ORGANISM: Aspergillus fumigatus
; US-10-128-714-5250

```

```

Query Match      3.1%; Score 52.4; DB 9; Length 3495;
Best Local Similarity 48.1%; Pred. No. 3.6e-05;
Matches 149; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

```

```

Qy 1173 GAATGCTTGGAAAGATTAACATTAAGCAATGCGATTCAAGTTCAGATTTGGGACAG 1232
Db 2223 GAACACTTGTGTGTGTCCCGTGAAGCCATGCGCGGAGTCCCGTCTCCATCA 2282
Qy 1233 GCTGCTGAGGACCAACGAGATGCTCTGAGCGGCTGAGTGGCTCTGATCTGCT 1292
Db 2283 AATACAGCGGGCCATTAAGAACTATCTGTAAGGCGGAAACGGGCTGCTCCGGACCA 2342
Qy 1293 GGGAGAGAGGCGTGGCCCTCTTGCAAGAACATCTGCACTCGGACCGACCGCGAG 1352
Db 2343 ACGGTTGCGGATTTGACCTGCGGTATGACCGGGTCTCTTACGATGACCAAG 2402
Qy 1353 CAGAGGCTCTCCATGGGAGAAAAGGCTATGAGAGGGTGAAGAAATGTTCTAGAGAC 1412
Db 2403 GACTTGACCGCATGTCAGCGGTGAGCAAGCGAGGTGAGAGGAGGAGTTCGTGCT 2462
Qy 1413 CACATGCTGAGAGATGCGCGCGGTGTAAGAGATGCTCGAGAAATACAGAGAC 1472
Db 2463 GCAATGGGTACAGGCTGAGAGAGAGATCCGATATGCTCAGCAAGAAAGACGGCG 2522
Qy 1473 TTCAGGTCTT 1482
Db 2523 TTCAGCGTT 2532

```

```

RESULT 5
US-10-156-761-3022
; Sequence 3022, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3022
; LENGTH: 1326
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1326)

```

```

US-10-156-761-3022

```

```

Query Match      2.9%; Score 49.8; DB 9; Length 1326;
Best Local Similarity 52.7%; Pred. No. 0.00012;
Matches 108; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

```

```

Qy 56 ACCGGCATGCGAAAGACCCCGTCTTGGCGGTGCGCGGTGCGCGAGCCCGCGGCC 115
Db 261 ACGGGGTGCGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 320
Qy 116 GATTCAACACGACCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 175
Db 321 CGCGGCGACCCGACCATACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380
Qy 176 CACAGAGAGGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 235
Db 381 CTTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 440
Qy 236 CGCGGCTCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 260
Db 441 CGGCTCTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 465

```

```

RESULT 6
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
; US-10-156-761-1

```

```

Query Match      2.9%; Score 49.8; DB 9; Length 9025608;
Best Local Similarity 52.7%; Pred. No. 0.079;
Matches 108; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

```

```

Qy 56 ACCGGCATGCGAAAGACCCCGTCTTGGCGGTGCGCGGTGCGCGAGCCCGCGGCC 115
Db 3794466 ACGGGGTGCGCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3794407
Qy 116 GATTCAACACGACCCAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 175
Db 3794406 CGCGGCGACCCGACCATACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3794347
Qy 176 CACAGAGAGGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 235
Db 3794346 CTTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 3794287
Qy 236 CGCGGCTCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 260
Db 3794286 CGGCTCTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3794262

```

```

US-08-902-572-13
: Sequence 13, Application US/08902572
: Patent No. US20020068706A1
: GENERAL INFORMATION:
: APPLICANT: Gyuris, Jeno
: APPLICANT: Lamphere, Lou
: APPLICANT: Beach, David H.
: TITLE OF INVENTION: INHIBITORS OF CELL-CYCLE PROGRESSION AND
: TITLE OF INVENTION: RELATED THERETO
: NUMBER OF SEQUENCES: 34
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: FOLEY, HOAG & ELLIOT LLP
: STREET: One Post Office Square
: CITY: Boston
: STATE: MA
: COUNTRY: USA
: ZIP: 02109-2170
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/902,572
: FILING DATE: 29-JUL-1997
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Vincent, Matthew P.
: REGISTRATION NUMBER: 36,709
: REFERENCE/DOCKET NUMBER: MIV-069, 03
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-832-1000
: TELEFAX: 617-832-7000
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 912 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA.
:
: US-08-902-572-13
:
: Query Match 2.88; Score 48.2; DB 8; Length 912;
: Best Local Similarity 47.28; Pred. No. 0.0003;
: Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0.
:
: QY 3 TCGGCTCGAGTTTATCCGAGCCGACAGTCTCTCTCGGAGCCGACGCGTCCGACGGCG 62
: Db 325 TCCGAGAGGGGCGGAGCGCACACCAACGACCGCCCCCGGGGCCCCCGAACCAGGGGTS 384
:
: QY 63 ATGGCAAGAACCCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 122
: Db 385 GCACATCAGGCCCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 444
:
: QY 123 AACGGAGCCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 182
: Db 445 CAGAAATCCGCGGACATCCGACAGCCCGCGCTCTGACGGGCGCAACCGGATCCAGAC 504
:
: QY 183 GGGTCTTCTCTCGCGGTGCTCTGCGAGACCTTTCGACAGGCGTGGAGACCCGCGGCC 242
: Db 505 CCGGCGCAGGGGCTGGCGAGAAAGCTGACATTTCAGACGCGCCCCCGAACCAGGAGG 564
:
: QY 243 CTCGAACCGCGCTGCGCAGAGCGGAGTTCCTCGGCTTCAGAGTTCAGAGTCTGTCCTC 302
: Db 565 CAGTGAACCCCCCGGGGCGGCTTTTAAAGAGCGGTCTTTCGCGCGCGGATCGGCGCG 624
:
: QY 303 CTCGCTCC 311
: Db 625 CTGGCGGCC 633

```

```

RESULT 8
US-10-161-499-11
Sequence 11, Application US/10161499
Publication No. US20030044427A1
GENERAL INFORMATION:
APPLICANT: Howley, Peter M.
APPLICANT: Benson, John
APPLICANT: Kasukawa, Hiroaki
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATING
FILE REFERENCE: HMV-041.01
FILE REFERENCE: PAPILOMAVIRUS-INFECTED CELLS
CURRENT APPLICATION NUMBER: US/10/161,499
CURRENT FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: US/09/347,504
PRIOR FILING DATE: 1999-07-02
NUMBER OF SEQ ID NOS: 79
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 912
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Ndel-BclRI
US-10-161-499-11

Query Match                2.8%; Score 48.2; DB 9; Length 912;
Best Local Similarity 47.2%; Pred. No. 0.0003;
Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

Oy      3 TCGGCTGAGATTGATTCGGAAGCCCAAGTCTCTCCGTGGGCGCCAGCGGTCGACCGGG 62
Db      325 TCGGAGAGGCGCGGAGACGACACCCACACCGCCCGGGCGCCCGGAGAACCAAGGGGT 384

Oy      63 ATGCAAAAGCCCGCTGCTTGGGTGGGTGGCGGCTGCGCGGAGCCCGGGCGGCTTAC 122
Db      385 GCGACTAAGGCCCGCGCGGCGCCGGGGGAGAGACACCGCGGAGAAATCGGCCAG 444

Oy      123 AACCGAGCCAGCTCTCTCTCTCTCTCTCTGAGCGCTGCGACCTTCGACACAGCA 182
Db      445 CGAAGATCGCGCGCACTCCAGAGCGCCCGCGGTGAGCGGCGCGAACCGATCCAGACA 504

Oy      189 GGGTCTCTCTCGCGGGTGGCTTCGAGACCTTGGCAGCGCGGTGGGAGCCCGCGCC 242
Db      505 CCCGCGAGGGGTGGCCAGAAAGCTGACCTTATGACCGCCCGCCCAAAACCCGAGCG 564

Oy      243 CTCACACCGCGCTGCGCCAGCGGAGAGTCCCTCGGGTTCAATGAGTCCAACTGTC 302
Db      565 CCATGACCGCCCGGGGTGGCGGCTTACAAAGCGCGTTCGCGCGCGGTGGGCGC 624

Oy      303 CTCGTCTCC 311
Db      625 CTGGCGGCGC 633

RESULT 9
US-09-862-179A-40
Sequence 40, Application US/09862179A
Patent No. US20020147306A1
GENERAL INFORMATION:
APPLICANT: Lin, Danny
APPLICANT: Pearson, Anthony
TITLE OF INVENTION: PEPTIDES THAT MODULATE THE INTERACTION OF B CLASS EPHRINS
FILE OF INVENTION: AND PDZ DOMAINS
FILE REFERENCE: MTSI-P01-009
CURRENT APPLICATION NUMBER: US/09/862,179A
CURRENT FILING DATE: 2001-05-21
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn version 3.1
SEQ ID NO 40
LENGTH: 912
TYPE: DNA

```



ORGANISM: HSV-1  
US-09-862-179A-40

Query Match 2.8%; Score 48.2; DB 10; Length 912;  
Best Local Similarity 47.2%; Pred. No. 0.0003;  
Matches 146; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

Qy 3 TCGGCTGAGTTGATCCGAGCCGACAGTCTCTCTCGAGCCGACCGCGTCCGACCGCGG 62  
Db 325 TCCGAGAGGCGCGAGACACCCACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 384  
Qy 63 ATGCAAAAGACCCCGTCTGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTTAC 122  
Db 385 GCGACTAAGCG 444  
Qy 123 AACGGAACCGAGCT 182  
Db 445 CCAGATTCG 504  
Qy 183 GGGTTCT 242  
Db 505 CCGCGGAGGCG 564  
Qy 243 CTCAACACCG 302  
Db 565 CCATGAGACCG 624  
Qy 303 CTCGCTCC 311  
Db 625 CTGGGGGC 633

## RESULT 10

US-10-156-761-5655

Sequence 5655, Application US/10156761  
Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 5655

LENGTH: 918

TYPE: DNA

ORGANISM: Streptomyces avermitilis

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(918)

US-10-156-761-5655

## Query Match

2.8%; Score 48; DB 9; Length 918;  
Best Local Similarity 50.0%; Pred. No. 0.00034;  
Matches 120; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

Qy 73 CCGCGTCTGCTGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 132  
Db 566 CCGGATCCCGGAGACCGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 625  
Qy 133 AGGTCT 192  
Db 626 AACGCTCGCGCTTCCGCTGTCCTCCGCTTACGCTTCCGCGCGCGCGCGCGCGCGCG 685

Qy 193 TCCGAGTCCCTCGAGACCTTTGCAAGCGCGTGGGAGACCCCGCGCGCGCGCGCGCG 252  
Db 686 CCGCGTGGCG 745  
Qy 253 CCGTCCGAGCGGAGATCCCTCGGTTTATGAGTCCAAAGTCTGCTCTCTCTCTCC 312  
Db 746 CCG 805

## RESULT 11

US-10-156-761-1

Sequence 1, Application US/10156761  
Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 1

LENGTH: 9025608

TYPE: DNA

ORGANISM: Streptomyces avermitilis

FEATURE:

NAME/KEY: misc feature

LOCATION: (4187715)

OTHER INFORMATION: a, t, c, g, other or unknown

US-10-156-761-1

## Query Match

2.8%; Score 48; DB 9; Length 9025608;  
Best Local Similarity 50.0%; Pred. No. 0.3;  
Matches 120; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

Qy 73 CCGCGTCTGCTGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 132  
Db 6860939 CCGGATCCCGGAGACCGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6860998  
Qy 133 AGGTCT 192  
Db 6860999 AACGCTCGCGCTTCCGCTGTCCTCCGCTTACGCTTCCGCGCGCGCGCGCG 6861058  
Qy 193 TCCGAGTCCCTCGAGACCTTTGCAAGCGCGTGGGAGACCCCGCGCGCGCGCGCG 252  
Db 6861059 CCGGATCCCGGAGACCGGAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6861118  
Qy 253 CCGTCCGAGCGGAGATCCCTCGGTTTATGAGTCCAAAGTCTGCTCTCTCTCTCC 312  
Db 6861119 CCG 6861178

## RESULT 12

US-10-184-644-332/C

Sequence 332, Application US/10184644  
Publication No. US20030044930A1

GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.

APPLICANT: Chen, Jiam

APPLICANT: Desnoyers, Luc

APPLICANT: Goddard, Audrey

APPLICANT: Godowski, Paul J.

APPLICANT: Gurney, Austin L.

Query Match 2.8%; Score 47; DB 9; Length 520;  
Best Local Similarity 18.6%; Pred. No. 0.00047;  
Matches 80; Conservative 89; Mismatches 262; Indels 0; Gaps 0

RESULT 13  
US-10-184-634-332/C

Sequence 332, Application US/10184634  
Publication No. US20030068684A1  
GENERAL INFORMATION:  
APPLICANT: Baker, Kevin P.  
APPLICANT: Chen, Jian  
APPLICANT: Denoyers, Luc  
APPLICANT: Goddard, Audrey  
APPLICANT: Godowski, Paul J.  
APPLICANT: Gurney, Austin L.  
APPLICANT: Pan, James  
APPLICANT: Smith, Victoria  
APPLICANT: Watanabe, Colin K.  
APPLICANT: Wood, William I.  
APPLICANT: Zhang, Zemin  
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
TITLE OF INVENTION: ACIDS ENCODING THE SAME

Query Match 2.8%; Score 47; DB 9; Length 520;  
Best Local Similarity 18.6%; Pred. No. 0.00047;  
Matches 80; Conservative 89; Mismatches 262; Indels 0; Gaps 0

RESULT 14  
US-10-156-761-1894

```

Sequence 1894, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: APPLICANT: IKEYA, HARUO
APPLICANT: ISHIZAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT FILING NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1894
LENGTH: 1038
TYPE: DNA
ORGANISM: Streptomyces avermitilis

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 18:55:10 ; Search time 1706 Seconds  
(without alignments)  
16157.514 Million cell updates/sec

Title: US-09-938-294-2

Perfect score: 1702

Sequence: 1 atccgcctcagcttgcaccc.....gagagatcctcaatcttagt 1702

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estcom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_hiv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vtc:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_oher:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	815.4	47.9	843	11	AY107383
2	615	36.1	615	11	AY108664
3	612	36.0	613	10	AM067129
4	551	32.4	721	12	BC464581
5	519.6	30.5	721	12	BR628089
6	512	30.1	670	14	BQ788642

7	502.6	29.5	557	12	BC464257	BC464257 EML 71 E0
8	492.2	28.9	497	12	BC268271	BC268271 1000172A0
9	417	24.5	627	13	B1959909	B1959909 HVSnm002
10	401.6	23.6	439	12	BG560599	BG560599 RH122 59
11	395	23.2	685	14	BQ48371	BQ48371 NF067E07F
12	375.2	22.0	641	13	B1271361	B1271361 NF050A09F
13	360.4	21.2	643	10	AW186747	AW186747 BNLH1137
14	342.4	20.1	375	10	AW257867	AW257867 687063C09
15	342.2	19.2	437	14	BQ464581	BQ464581 HF02101F
16	325.4	19.1	551	13	B1421630	B1421630 EST532296
17	314.4	18.5	513	10	BR402653	BR402653 CSB010B11
18	314.4	18.5	513	14	BQ608401	BQ608401 BRY 4306
19	307	18.0	543	14	BQ696254	BQ696254 san89a12
20	302.8	17.8	443	14	C74728	C74728 C74729 Rice
21	218.4	12.8	461	10	BE209946	BE209946 sc36h10.Y
22	215.8	12.7	390	10	BE022039	BE022039 sm65906.Y
23	211.2	12.4	396	13	BJ208273	BJ208273 BJ208273
24	203.2	11.9	563	12	BF270892	BF270892 GA_EB000
25	203.2	11.9	711	10	AM266508	AM266508 L0-1363T3
26	202.6	11.9	782	13	B1971272	B1971272 GM830013A
27	185.2	10.9	425	14	T76431	T76431 11209 Lambd
28	185	10.9	360	9	AJ473497	AJ473497 AJ473497
29	185	10.9	360	9	AJ473498	AJ473498 AJ473498
30	171.2	10.1	863	17	BH501249	BH501249 BGVX70TF
31	160.4	9.4	715	17	B78162	B78162 T31117TF TA
32	158.2	9.3	553	10	BE472046	BE472046 EST416899
33	154	9.0	634	13	B173349	B173349 B173349
34	150.2	8.8	251	13	B1121818	B1121818 F047271F
35	148.6	8.7	647	12	BF642258	BF642258 NF066A10I
36	135.8	8.0	276	10	AV540026	AV540026 AV540026
37	134.2	7.9	724	17	B21743	B21743 T31117-T7 T
38	131.6	7.7	669	13	BJ171310	BJ171310 B171310
39	129.6	7.6	629	9	AU239738	AU239738 AU239738
40	125.8	7.4	288	17	BH810193	BH810193 SALK_0442
41	125.4	7.4	435	14	BQ662962	BQ662962 HS04P17u
42	104.6	6.1	569	13	B1267795	B1267795 NF11D091
43	102.2	6.0	688	14	BQ114748	BQ114748 EST600324
44	101.6	6.0	707	14	BQ118269	BQ118269 EST603845
45	101.6	6.0	714	14	BQ118358	BQ118358 EST603934

## ALIGNMENTS

RESULT 1  
LOCUS AY107383 843 bp mRNA linear HTC 25-MAY-2002  
DEFINITION Zea mays PC0096987 mRNA sequence.  
ACCESSION AY107383  
VERSION AY107383.1 GI:21210461

KEYWORDS  
SOURCE  
ORGANISM

Zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
Clade; Panicoideae; Andropogoneae; Zea.

REFERENCE  
AUTHORS Hayney,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitesitt,M.S.,

TITLE  
Hathur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.

Maize Mapping Project/DuPont Consensus Sequences for Design of  
Overgo Probes

JOURNAL  
REFERENCE  
AUTHORS

Unpublished (2002)  
2 (bases 1 to 843)  
Coe,E.C.

Direct Submission  
Submitted (25-Apr-2002) Maize Mapping Project, University of  
Missouri, Columbia, MO 65211, USA

FEATURES  
source  
Location/Qualifiers

1..843  
/organism="Zea mays"  
/db\_xref="MaizeDB:635540"  
/db\_xref="taxon:4577"  
/clone="PC0096987"

/clone.lib="Maize Mapping Project/DuPont Cornsensus  
library"  
/note="this sequence is part of a project of EST  
assemblies resulting from the application of public  
contigs to seed Dupont contigs; this resource was  
assembled by DuPont as part of a collaboration for the  
overgo addressing of BACs in conjunction with the Maize  
Mapping Project"

BASE COUNT 184 a 231 c 224 g 199 t 5 others

Query Match 47.9% Score 815.4; DB 11; Length 843;  
Best Local Similarity 99.2% Pred. No. 5.5e-200;  
Matches 827; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

QY 19 CCGAGCCCAACAGTCTCTCTCTGCGGCCACCGGCTCCGACCGGAGTGCAGAAAGACCCCGT 78  
DB 10 CGAGGCCCAACAGTCTCTCTGCGGCCACCGGCTCCGACCGGAGTGCAGAAAGACCCCGT 69  
QY 79 CGTTGGGGTGG 138  
DB 70 CGTTGGGGTGG 129  
QY 139 TCT 198  
DB 130 TCT 189  
QY 199 GTGCGCTGCGAGACCTTGTGCGAGCGGCGGTGGGAGACCGCGGCGCTCAACCGCGCGTGG 258  
DB 190 GTGCGCTGCGAGACCTTGTGCGAGCGGCGGTGGGAGACCGCGGCGCTCAACCGCGCGTGG 249  
QY 259 CCGAGCGGAGTCCCTCGGGTTCTATAGAGTCCAGCTGCTGCTCTCTCTCTCTCTCTCTCT 318  
DB 250 CCGAGCGGAGTCCCTCGGGTTCTATAGAGTCCAGCTGCTGCTCTCTCTCTCTCTCTCTCT 309  
QY 319 TCT 378  
DB 310 TCT 369  
QY 379 GCTTCGCAAGTGTGTGATTAACAACAGAGTTCACAAAGAAATGATGTCAATATTA 438  
DB 370 GCTTCGCAAGTGTGTGATTAACAACAGAGTTCACAAAGAAATGATGTCAATATTA 429  
QY 439 GCTTCGCAAGTGTGTGATTAACAACAGAGTTCACAAAGAAATGATGTCAATATTA 498  
DB 430 GCTTCGCAAGTGTGTGATTAACAACAGAGTTCACAAAGAAATGATGTCAATATTA 489  
QY 499 CAGTTGATTTGCTCTTAAAGCTGATCTGTATCTTAAACATGCTGTGTCTGCGAAAGT 558  
DB 490 CAGTTGATTTGCTCTTAAAGCTGATCTGTATCTTAAACATGCTGTGTCTGCGAAAGT 549  
QY 559 GGGTTGACCTGTGTGAAAGATCATGTTCTTAAAGTCTTCCGAGAGATTTGTGTGGA 618  
DB 550 GGGTTGACCTGTGTGAAAGATCATGTTCTTAAAGTCTTCCGAGAGATTTGTGTGGA 609  
QY 619 TCCATGAATATGCGGGGCTTACTTAAAGTGTGAATATGTCAAACATCTCTCTCTTGTG 678  
DB 610 TCCATGAATATGCGGGGCTTACTTAAAGTGTGAATATGTCAAACATCTCTCTCTTGTG 669  
QY 679 CTGAGACCATGATTTCTCATACACGGCTGAGTATTTGATATGACAGATGAGCATGCGATC 738  
DB 670 CTGAGACCATGATTTCTCATACACGGCTGAGTATTTGATATGACAGATGAGCATGCGATC 729  
QY 739 GCGTGAATAATACAGATGCAAACTTATGTTTCACTGCGGGAATGTGAATAAATA 798  
DB 730 GCGTGAATAATACAGATGCAAACTTATGTTTCACTGCGGGAATGTGAATAAATA 789  
QY 799 TGAAGTTGCTGAAGACAAT-GTGGCAAGAGAGTCTTACGGGAACATATTCGT 851  
DB 790 TGAAGTTGCTGAAGACAAGTGTGGCAAGAGAGTCTTACGGGAACATATTCGT 843

RESULT 2

AY108664  
LOCUS AY108664 615 bp mRNA linear HTC 25-MAY-2002  
DEFINITION Zea mays PC0096986 mRNA sequence.  
ACCESSION AY108664  
VERSION AY108664.1 GI:21211836  
KEYWORDS HTC.  
SOURCE Zea mays.  
ORGANISM Zea mays.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE  
AUTHORS Hainey, C.F., Dolan, M., Miao, G.H., Vogel, J.M., Whitlitt, M.S.,  
Arthur, L.W., Hanafey, M., Morgante, M., and Tingey, S.V.  
Maize Mapping Project/DuPont Consensus Sequences for Design of  
Overgo Probes  
Unpublished (2002)  
2 (bases 1 to 615)  
Coe, E.C.  
Direct Submission  
Submitted (25-APR-2002) Maize Mapping Project, University of  
Missouri, Columbia, MO 65211, USA  
Location/Qualifiers  
1. 615  
/organism="Zea mays"  
/db\_xref="taxon:4577"  
/clone="PC0096986"  
/clone.lib="Maize Mapping Project/DuPont Cornsensus  
library"  
/note="this sequence is part of a project of EST  
assemblies resulting from the application of public  
contigs to seed Dupont contigs; this resource was  
assembled by DuPont as part of a collaboration for the  
overgo addressing of BACs in conjunction with the Maize  
Mapping Project"

BASE COUNT 174 a 123 c 164 g 154 t  
ORIGIN

Query Match 36.1% Score 615; DB 11; Length 615;  
Best Local Similarity 100.0% Pred. No. 3.6e-148;  
Matches 615; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 721 ATAGAGAGACATGATGCGCTGAAATATACAGATGCCAAACTTATGTTGTCACCTGG 780  
DB 1 ATAGAGAGACATGATGCGCTGAAATATACAGATGCCAAACTTATGTTGTCACCTGG 60  
QY 781 GGAATAGTAAGAATCTAATGAGAGTGTCTGAGACATGTGCAAGAGATCTTACGGG 840  
DB 61 GGAATAGTAAGAATCTAATGAGAGTGTCTGAGACATGTGCAAGAGATCTTACGGG 120  
QY 841 AACATATTCGTGAATCCCTTGAAGTACGAGTGAAGATCTCTGTTGCAATTAACA 900  
DB 121 AACATATTCGTGAATCCCTTGAAGTACGAGTGAAGATCTCTGTTGCAATTAACA 180  
QY 901 GTGTATCACAGGAAGGAGCAAGATTTATTTTCAAGCATTTTACAGCTTTCAGC 960  
DB 181 GTGTATCACAGGAAGGAGCAAGATTTATTTTCAAGCATTTTACAGCTTTCAGC 240  
QY 961 TCATCCAAACAGGAATCTTAAAGTGTCTGATATCATGCTGTATGTTGGAAGATG 1020  
DB 241 TCATCCAAACAGGAATCTTAAAGTGTCTGATATCATGCTGTATGTTGGAAGATG 300  
QY 1021 TTAAATGCTCAGAACCAATTTGAGACTCAGTTAGTGAATTTGTGTGAAGACAGATTC 1080  
DB 301 TTAAATGCTCAGAACCAATTTGAGACTCAGTTAGTGAATTTGTGTGAAGACAGATTC 360  
QY 1081 ATGACCGTGTCAATTTTGTGAACAAGACATTTGCAAGTGGCCCTTACTTGGCAGCAATTG 1140  
DB 361 ATGACCGTGTCAATTTTGTGAACAAGACATTTGCAAGTGGCCCTTACTTGGCAGCAATTG 420  
QY 1141 ATGTCCTTTGCAATTTCTCAGGGCCGTGAGAAATCTTTGGAAGATTAACAATTGAAG 1200

Db 421 ATGTCTTGTTCAGAAATTCAGGGCCGTGGAGAAATGCTTGGAGAGATTAACAATTGAAG 480

Qy 1201 CAATGGCATTCAGTTGTCAGATTAATTTGGGCAAGGCTGTGGAGGAGACCAAGGATGCTCC 1260

Db 481 CAATGGCATTCAGTTGTCAGATTAATTTGGGCAAGGCTGTGGAGGAGACCAAGGATGCTCC 540

Qy 1261 TGAACGGCTGACTGCTGCTTCTGCAATCTCTCTGGAGAGAGAGGCTGCTCTTTGCA 1320

Db 541 TGAACGGCTGACTGCTGCTTCTGCAATCTCTCTGGAGAGAGAGGCTGCTCTTTGCA 600

Qy 1321 AGAATCATCTCAGAC 1335

Db 601 AGAATCATCTCAGAC 615

RESULT 3

AM067129/c 633 bp mRNA linear EST 12-OCT-1999

LOCUS 683016A01.x1 683 - 14 day immature embryo from Hake lab (HS) Zea

DEFINITION mays cDNA, mRNA sequence.

ACCESSION AM067129

VERSION AM067129.1 GI:6022201

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays

REFERENCE Bukariyola; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Zea.

REFERENCE 1 (bases 1 to 633)

AUTHORS Walbot V

TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University

JOURNAL Unpublished (1999)

COMMENT Contact: Walbot V

Department of Biological Sciences

Stanford University

855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 683016 row: A column: 01.

Location/Qualifiers

source

1. 633

/organism="Zea mays"

/cultivar="973"

/db\_xref="taxon:4577"

/clone\_lib="683 - 14 day immature embryo from Hake lab (HS)"

/issue\_type="embryo"

/dev\_stage="14 days after pollination"

/lab\_host="DH10B"

/note="Organ: embryo; Vector: pBKCMV (Stratagene's Zap Express); Site 1: XhoI; Site 2: EcoRI; Directionally cloned, 14 day immature embryo library created with Stratagene's Zap Express cDNA protocol."

BASE COUNT 161 a 174 c 134 g 159 t 5 others

ORIGIN

Query Match 36.0%; Score 612; DB 10; Length 633;

Best Local Similarity 99.2%; Pred. No. 2.2e-147;

Matches 612; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Db 1079 TCATGACCGGTGCATTGTGAGACAGACATTCGAGGCGCTTACTTGCGAGCAAT 1138

Db 633 TCATGACCGGTGCATTGTGAGACAGACATTCGAGGCGCTTACTTGCGAGCAAT 574

Qy 1139 TCATGCGTTGTCAGAAATTCAGAGCGCTGAGAAATGCTTTGAAAGATTAACAATTGA 1198

Db 573 TCATGCGTTGTCAGAAATTCAGAGCGCTGAGAAATGCTTTGAAAGATTAACAATTGA 514

Qy 1199 AGCAATGCAATTCAGTTGTCAGATTAATTTGGGCAAGGCTGTGGAGGAGACCAAGGATGCT 1258

Db 513 AGCAATGCAATTCAGTTGTCAGATTAATTTGGGCAAGGCTGTGGAGGAGACCAAGGATGCT 454

Qy 1259 CTTGAGCGGCTGACGTGACCTTCTGCACTCTGCTGGAGAGAGGCGTGGCCCTCTTTC 1318

Db 453 CTTGAGCGGCTGACGTGACCTTCTGCACTCTGCTGGAGAGAGGCGTGGCCCTCTTTC 394

Qy 1319 AAAAATCATCTGCACTCTGCAAGCCAGCCAGACAGAGGCTTCCATGGGGGAAAAGG 1378

Db 393 AAAAATCATCTGCACTCTGCAAGCCAGCCAGACAGAGGCTTCCATGGGGGAAAAGG 334

Qy 1379 CTATGCGAGGTGAGAGAAATGTTCAATGAGACACATGCTGAGAGAGATCGGGCGT 1438

Db 333 CTATGCGAGGTGAGAGAAATGTTCAATGAGACACATGCTGAGAGAGATCGGGCGT 274

Qy 1439 GTTGAAGATGCTGCTGAGAGAAATCAAGAGACTCCAGGCTTTCGCTGGCCGCC 1498

Db 273 GTTGAAGATGCTGCTGAGAGAAATCAAGAGACTCCAGGCTTTCGCTGGCCGCC 214

Qy 1499 ATCAGCTTCGCTTAACTATGTTGAATGTTTACGGGCTACGCTGAGGCT 1558

Db 213 ATCAGCTTCGCTTAACTATGTTGAATGTTTACGGGCTACGCTGAGGCT 154

Qy 1559 GTAACCTGATGATGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1618

Db 153 GTAACCTGATGATGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 94

Qy 1619 CCATGTCGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 1678

Db 93 CCATGTCGATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTC 34

Qy 1679 GATTGAGGAGATTTC 1695

Db 33 GATTGAGGAGATTTC 17

RESULT 4

BG464581 678 bp mRNA linear EST 20-MAR-2001

LOCUS BG464581

DEFINITION BM1 71 B08.g1\_A002 Embryo 1 (BM1) Sorghum bicolor cDNA, mRNA

ACCESSION BG464581

VERSION BG464581.1 GI:13393230

KEYWORDS EST.

SOURCE sorghum.

ORGANISM Sorghum bicolor

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Sorghum.

REFERENCE 1 (bases 1 to 678)

AUTHORS Reid, S.P., Cordomier-Pratt, M.-M., Gingle, A. and Pratt, L.H.

JOURNAL An EST database from Sorghum: developing embryos

COMMENT Unpublished (2000)

Contact: Cordomier-Pratt MM

Laboratory for Genomics and Bioinformatics

The University of Georgia, Department of Plant Biology

Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

Tel: 706 542 1860

Fax: 706 583 0210

Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions below Pined quality 16. The threshold for highest quality sequence is 20.

Seq primer: PolyTmix

High quality sequence start: 26

High quality sequence stop: 669

POLYANo.

Location/Qualifiers

source

1. 678

/organism="Sorghum bicolor"

/db\_xref="taxon:4558"

/clone\_lib="Embryo 1 (BM1)"

/note="Organ: Embryos germinated for 24 hr; Vector: pLuescript II from Lambda Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made from poly-A RNA in the cloning

vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 186 a 135 c 191 g 164 t 2 others

ORIGIN

Query Match 32.4%; Score 551; DB 12; Length 678;  
Best Local Similarity 95.0%; Pred. No. 1.3e-11;  
Matches 569; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 890 AATAAATAACAGTGTATCAGAGAAAGGACAAAGCTTATTTCTTCAAGATTTTATCA 949  
DB 1 AATAAATAACAGTGTATCAGAGAAAGGACAAAGCTTATTTCTTCAAGATTTTATCA 60  
QY 950 GGGTTTGACGTCACTCCACAGCAAGAACTTAAAGTCCCTAGAAATACATGCTAGATTGT 1009  
DB 61 GGGTTTGACGTCACTCCACAGCAAGAACTTAAAGTCCCTAGAAATACATGCTAGATTGT 120  
QY 1010 GGGAACTGATGTAAATGCTCAGACCAAAATTTGACATCACTTACGTAAGCTTTGGTGA 1069  
DB 121 GGGAACTGATGTAAATGCTCAGACCAAAATTTGACATCACTTACGTAAGCTTTGGTGA 180  
QY 1070 GAAACAGATTCAATGACGCTGCTCATTGTCGAAACAAATTTGGCAGTGGCCCTTACTT 1129  
DB 181 GAAATGATTCATGACGCTGCTCATTGTCGAAACAAATTTGGCAGTGGCCCTTACTT 240  
QY 1130 GGGAGCAATTGATGCTGCTGTCGAAATTTCCAGAGGCGCTGAGAAATGCTTTGGAAGAT 1189  
DB 241 GGGAGCAATTGATGCTGCTGTCGAAATTTCCAGAGGCGCTGAGAAATGCTTTGGAAGAT 300  
QY 1190 AACAAATTTGAAAGCAATGCAATTCAGATTTGCGGACCGCTGCTGAGAGGACAC 1249  
DB 301 AACAAATTTGAAAGCAATGCAATTCAGATTTGCGGACCGCTGCTGAGAGGACAC 360  
QY 1250 GGAATGCTGCTGAGAGGCTGCACTGCGCTTCTGATCTGCTGCGGAAGAGAGGCGTGGC 1309  
DB 361 GGAATGCTGCTGAGAGGCTGCACTGCGCTTCTGATCTGCTGCGGAAGAGAGGCGTGGC 420  
QY 1310 GCTCTTTGCAAGAAACATGCTGCACTGCGCAAGCCGAGCAGAGAGGCTTCCATGGG 1369  
DB 421 GCTCTTTGCAAGAAACATGCTGCACTGCGCAAGCCGAGCAGAGAGGCTTCCATGGG 480  
QY 1370 GGAAGAGGCTATGCGAGGCTGGAAGAAATGTCATGAGACCAATGAGGCTGAGAGAT 1429  
DB 481 GGAAGAGGCTATGCGAGGCTGGAAGAAATGTCATGAGACCAATGAGGCTGAGAGAT 540  
QY 1430 CGCGGCGGTGTGAAGATGCTCTGAGAAATCAGAGAGCACTCAAGTCTTGAAGCTT 1488  
DB 541 CGCGGCGGTGTGAAGATGCTCTGAGAAATCAGAGAGCACTCAAGTCTTGAAGCTT 599

RESULT 5  
BF629089 721 bp mRNA linear EST 22-OCT-2001  
LOCUS BF629089  
DEFINITION HVSMDB0009N03f Hordeum vulgare seedling shoot EST library  
HVCNDA0002 (Dehydration stress) Hordeum vulgare cDNA clone  
HVSMB0009N03f, mRNA sequence.  
BF629089  
ACCESSION BF629089.2 GI:13091072  
VERSION EST  
KEYWORDS Hordeum vulgare.  
SOURCE Hordeum vulgare.  
ORGANISM Hordeum vulgare.  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae; Triticeae; Hordeum.  
1 (bases 1 to 721)  
Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Fritsch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simons, J., Choi, D.W., Fenton, R.D., Oates, R. and Main, D.  
Development of a genetically and physically anchored EST resource for barley genomes: Morex drought-stressed seedling shoot cDNA library  
Unpublished (2001)  
COMMENT On Dec 19, 2000 this sequence version replaced gi:11893247.

Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Total hg bases = 588  
Seq primer: AATTACCTCCTCAATGAGG  
High quality sequence stop: 716.  
Location/Qualifiers  
1. 721  
/organism="Hordeum vulgare"  
/cultivar="Morex"  
/db\_xref="taxon:4513"  
/clone="HVSMDB0009N03f"  
/clone\_1fb="Hordeum vulgare seedling shoot EST library  
HVCNDA0002 (Dehydration stress)"  
/issue\_type="seedling shoot"  
/lab\_host="TUC121"  
/note="Vector: lambdaZAP; Site\_1: EcoRI; Site\_2: XhoI;  
Seeds were surface sterilized then germinated under axenic  
conditions in the dark at room temperature on filter paper  
with water, mycstatin and cefotaxime in covered  
crystallization dishes. Five-day old seedlings were  
incubated at 90% RH for 24 hr. Shoots were then harvested,  
total RNA was prepared, poly(A) RNA was purified, one  
primary unamplified cDNA library was made, 600000 pfu were  
in vivo excised to give pluescript SK(-) cDNA phagemids.  
These steps were performed in the TJ Close laboratory at  
the University of California, Riverside (Choi, Close,  
Fenton). Phagemids were plated and picked at the Clemson  
University Genomics Institute (CUGI) (Begum, Palmer,  
Fritsch, Atkins and Wing). Plasmid DNA preparations, DNA  
sequencing and sequence analysis were performed at CUGI  
(Wing, Yu, Fritsch, Henry, Simons, Oates, Rambo, Main).  
The sequence has been trimmed to remove vector sequence  
and contains a minimum of 100 bases of phred value 20 or  
above. For more details on library preparation and  
sequence analysis see  
http://www.genome.clemson.edu/projects/barley. To order  
this clone see http://www.genome.clemson.edu/orders Also  
see Close TJ, Wing R, Kleinhofs A, Wise R (2001)  
Genetically and physically anchored EST resources for  
barley genomes. Barley Genetics Newsletter 31:29-30.  
(http://wheat.pw.usda.gov/gspages/bgm/31/cover.html)"

BASE COUNT 195 a 151 c 203 g 172 t

ORIGIN

Query Match 30.5%; Score 519.6; DB 12; Length 721;  
Best Local Similarity 83.7%; Pred. No. 1.8e-123;  
Matches 601; Conservative 0; Mismatches 114; Indels 3; Gaps 1;

QY 781 GGAATGATGAAGAACTAATGAAAGTTGCTGAGACCAATGTCGAAAGAGGCTTACGGG 840  
DB 1 GGAATGATGAAGAACTAATGAAAGTTGCTGAGACCAATGTCGAAAGAGGCTTACGGG 60  
QY 841 AACATATTGCTGAATCCCTTGGAGTACGAGATCCCTGTTTGCATTAATAACA 900  
DB 61 AACATATTGCTGAATCCCTTGGAGTACGAGATCCCTGTTTGCATTAATAACA 120  
QY 901 GTGTATCGAGGAAAGGACAAAGCTTATTTCTTCAAGATTTTATCAGGCTTTGGCAGC 960  
DB 121 GTGTATCGAGGAAAGGACAAAGCTTATTTCTTCAAGATTTTATCAGGCTTTGGCAGC 180  
QY 961 TCATCAACAGAGAACTTAAAGTCCCTAGAAATACATGCTGTAGTTGGAGAGTGATG 1020  
DB 181 TCATCAACAGAGAGTTAGAGAGGCCAAAGTGCATGCTGTAGTTGGAGAGTGATG 240  
QY 1021 TTAAATGCTCAGACCAATTTGAGACTGATGATGCTGCTTGTGTGAAGAACATTC 1080  
DB 241 TGAATGCTCAGACCAATTTGAGACTGATGATGCTGCTTGTGTGAAGAACATTC 300



QY 1081 ATGACCGTGTCCATTTTGTGAACAAGACATTTGGCAGTGGCCCTTACTTGGCAGCAATTG 1140  
 Db 301 ATGACCGTGTACCTTTGTCAACAAGACGTTGGCAGTGGCAGCTTATTTGGTGGCAATTG 360  
 QY 1141 ATGTCCTTTTTCAGAAATTTCTCAGGCGCTTGGAGAAATGCTTTGGAAAGATTAACAATTGAAG 1200  
 Db 361 ATGTCCTTTTTCAGAAATTTCTCAGGCGCTTGGAGAAATGCTTTGGAAAGATTAACAATTGAAG 420  
 QY 1201 CAATGGCATTCAAGTTCGACATTTGGGCGACCGCTGTCTGAGGAGCAGCAGATGCTCC 1260  
 Db 421 CAATGGCATTCAAGTTCGACATTTGGGCGACCGCTGTCTGAGGAGCAGCAGATGCTCC 480  
 QY 1261 TGGACGGCTGCACTGGCCCTTCTGATCTCTTGGGAGAGAGGCTTGGGCGCTTGGAA 1320  
 Db 481 TGGATGTTCAACCGGCTCTTCTGACCTCTGGAAGAGAGGCTGACGCTCTGCGCA 540  
 QY 1321 AGAATCATCTCAGATCTGCAAGCAGCCGAGCAGAGGCTCTCCATGGGAGAAAGGCT 1380  
 Db 541 GGAACATGTTAGGCTTGGAGGAGCAGATGTTAGCAGAGGAGTCTCCATGGGAAACAAGGCT 600  
 QY 1381 ATGGCAGGAGGAGAAATGTTCAATGAGCAGCAGATGCTGAGAGATGCGGCGGCT 1440  
 Db 601 ATGGCAGGAGGAGAGAGGTTCAATGAGCAGCAGATGCTGAGAGATGCGGCGGCT 660  
 QY 1441 TGAAGATGTCCT--GAGGAAATCAGAGAGCAGTCCAGGTCCTTGAAGCTTTGCCGTG 1495  
 Db 661 TGAAGAGAGTTCTTCAAGCAGTCTCGCAGCAGATCTCATCTTCTGAGCTTCATCATG 718

RESULT 6  
 BQ788642 670 bp mRNA linear EST 26-JUL-2002  
 LOCUS BQ788642

DEFINITION WHE4152\_B09\_C18ZS wheat CS whole plant cDNA library Triticum  
 aestivum cDNA clone WHE4152\_B09\_C18, mRNA sequence.

ACCESSION BQ788642  
 VERSION BQ788642.1 GI:21997114

KEYWORDS EST.

SOURCE bread wheat.

ORGANISM Triticum aestivum

REFERENCE Anderson, O.D., Akhunov, E., Zhao, S., Crossman, C., Deal, K., Dvorak, J.,  
 Lazo, G.R., Rausch, C.J., Wilson, C., and Woo, J.

AUTHORS The structure and function of the expressed portion of the wheat  
 genomes - Chinese Spring whole plant cDNA library

TITLE Unpublished (2002)

JOURNAL Contact: Olin Anderson

COMMENT US Department of Agriculture, Agriculture Research Service, Pacific  
 West Area, Western Regional Research Center  
 800 Buchanan Street, Albany, CA 94710, USA  
 Tel: 5105595773  
 Fax: 5105595818  
 Email: oanderson@w.usda.gov  
 Sequences have been trimmed to remove vector sequence and low  
 quality sequence with phred score less than 20  
 Seq primer: SK primer.

FEATURES  
 source

1. .670  
 /organism="Triticum aestivum"  
 /cultivar="Chinese Spring"  
 /db\_xref="taxon:4565"  
 /clone="WHE4152\_B09\_C18"  
 /clone\_lib="wheat CS whole plant cDNA library"  
 /tissue\_type="Roots, leaves, crown, stem and sheath"  
 /dev\_stage="Adult"  
 /lab\_host="E. coli SOLR"  
 /note="Vector: Lambda Uni-ZAP XR, excised phagemid  
 pluescript SK(-); Site 1: EcoRI; Site 2: XhoI; Plant  
 tissues from wheat cv. CS grown to full tillering stage in  
 greenhouse were collected at University of California,  
 Davis (Dan Dvorak's lab, K. Deal and E. Akhunov). Total

RNA was prepared from leaves (young leaf and third leaf),  
 whole roots, crown, stem and sheath tissues, and then  
 equal quantities of RNA were pooled from the these  
 samples. PolyA was purified from the pooled RNA, a cDNA  
 library was made, and the cDNA clones were in vivo excised  
 to give pluescript SK(-) phagemids in U. Dvorak's lab (E.  
 Akhunov, J. Dvorak) at the University of California,  
 Davis. Colony plating, plasmid DNA preparations and DNA  
 sequencing were performed in the OD Anderson lab (all  
 other authors)."

BASE COUNT 173 a 139 c 171 g 187 t

ORIGIN

Query Match 30.1%; Score 512; DB 14; Length 670;  
 Best Local Similarity 85.7%; Pred. No. 1.7e-12;  
 Matches 569; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 243 CTCACACCGCCGCTGCGCAGCGGAGTCCCTCGGTTTCATGAGTCCAGCTGCTC 302  
 Db 7 CACGAGGCTGGGGTGGCTGCGCTGAGCCCTCGGTTTCATGAGTCCAGCTGCTC 66

QY 303 CTGCTCTCCATGAGCTCTCCCTCTGAGTGTCACTTTTACTGATGGAATTAAGATT 362  
 Db 67 CTGCTCTCCATGAGCTCTCCCTCTGAGTGTCACTTTTACTGATGGAATTAAGATT 126

QY 363 CTTCTGAGGATGTTGGCTGCGCAGTGTGTGATTAACAACCAAGATCAACAAGAA 422  
 Db 127 CTTCTGAGGATGTTGGCTGCGCAGTGTGTGATTAACAACCAAGATCAACAAGAA 186

QY 423 AATGATGTCAATATATGCTTGGAGCAGATGATGATGATGATGATGATGATGAT 482  
 Db 187 AATGATGTCAATATATGCTTGGAGCAGATGATGATGATGATGATGATGATGAT 246

QY 483 GCTAGAGCAGAGAGGAGTGTATTTGCTTAAGCTGATGATGATGATGATGATGAT 542  
 Db 247 GCTAGAGCAGAGAGGAGTGTATTTGCTTAAGCTGATGATGATGATGATGATGAT 306

QY 543 GCGTGTGCTGCGAAGTGTGCTTGAACCTGTTGAAAGATGATGATGATGATGAT 602  
 Db 307 GCGTGTGCTGCGAAGTGTGCTTGAACCTGTTGAAAGATGATGATGATGATGATGAT 366

QY 603 AAGATTTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 662  
 Db 367 AAGATTTTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 426

QY 663 CATCTTCCTTTGTTGCTGAGGAGCAGATGATGATGATGATGATGATGATGAT 722  
 Db 427 CATCTTCCTTTGTTGCTGAGGAGCAGATGATGATGATGATGATGATGATGAT 486

QY 723 AGCAGACTAGGAGTGGCTGGAATAATACAGATGACCAAACTTATGTTGTTCACT 782  
 Db 487 AGCAGACTAGGAGTGGCTGGAATAATACAGATGACCAAACTTATGTTGTTCACT 546

QY 783 AATGATTAAGACTTAATGAAGTGTGTAAGCAATGTCGACAGAGAGTCTCGGGA 842  
 Db 547 AATGATTAAGACTTAATGAAGTGTGTAAGCAATGTCGACAGAGAGTCTCGGGA 606

QY 843 CATATTCGTAATCCCTTGAAGTACAGAGTGAAGATCTCTGTTGCAATTAATCA 902  
 Db 607 CATATTCGTAATCCCTTGAAGTACAGAGTGAAGATCTCTGTTGCAATTAATCA 666

QY 903 GTAT 906  
 Db 667 GTTT 670

RESULT 7  
 BQ464257 557 bp mRNA linear EST 20-MAR-2001  
 LOCUS BQ464257  
 DEFINITION EM1\_71\_E08\_b1\_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA  
 sequence.  
 ACCESSION BQ464257  
 VERSION BQ464257.1 GI:13392578



Db 257 AAGGCTATGCGAGGCTGAAGAAATGTTTCATGACCACTGGCTGAGAGGATCC 198  
 QY 1433 GCGGCTGTGAAGATGTCCTGAGAAATCAGAGAGACTCCAGCTTGTGAGCTTTGCC 1492  
 Db 197 GCGGCTGTGAAGATGTCCTGAATAAATACAGAGAGACTCCAGCTTGTGAGCTTTGCC 138  
 QY 1493 GTGGCCATCAGCTGGCTAAGCTGTAAGCTGATGATGAGGCTTACGCTGAGCTT 1552  
 Db 137 GTGGCCATCAGCTGGCTAAGCTGTAAGCTGATGATGAGGCTTACGCTGAGCTT 78  
 QY 1553 CAGGCTGTAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1612  
 Db 77 CAGGCTGTAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 18  
 QY 1613 ATTAGGCTATGCTGAT 1629  
 Db 17 ATTAGGCTATGCTGAT 1  
 RESULT 9  
 B1959909 627 bp mRNA linear EST 22-OCT-2001  
 LOCUS HVSMEN0022F11f Hordeum vulgare rachis EST library HVCDNA0015  
 DEFINITION (normal) Hordeum vulgare cDNA clone HVSMEN0022F11f, mRNA sequence.  
 ACCESSION B1959909  
 VERSION B1959909.1 GI:16311164  
 KEYWORDS EST.  
 SOURCE  
 ORGANISM Hordeum vulgare.  
 Hordeum vulgare.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 ; Triticeae; Hordeum.  
 1 (bases 1 to 627)  
 Wing, R., Close, T.J., Kleinbols, A., Wise, R., Chin, A., Begum, D.,  
 Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons,  
 J., Oates, R. and Main, D.  
 Development of a genetically and physically anchored EST resource  
 for barley genomes: Morex rachis cDNA library  
 Unpublished (2001)  
 Contact: Wing RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson, SC 29634, USA  
 Tel: 864 656 7288  
 Fax: 864 656 4293  
 Email: rwing@clemson.edu  
 Total hg bases = 508  
 Seq primer: ATTATCCCTCACTTAAGG  
 High quality sequence stop: 612.  
 Location/Qualifiers  
 1. 627  
 /organism="Hordeum vulgare"  
 /cultivar="Morex"  
 /db\_xref="taxon:4513"  
 /clone="HVSMEN0022F11f"  
 /clone\_11b="Hordeum vulgare rachis EST library HVCDNA0015  
 (normal)"  
 /rissue\_cye="Rachis"  
 /lab\_host="TUC121"  
 /note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:  
 XhoI; Plants were grown at Washington State University,  
 Pullman, WA in a greenhouse, the rachises were excised and  
 frozen in liquid nitrogen (Kleinbols lab). In the TUC121  
 lab at the University of California, Riverside total RNA  
 was prepared, poly(A) was purified, one primary  
 unamplified cDNA library was made, and 1 million pfu were  
 in vivo excised to give pBluescript SK(-) cDNA phagemids  
 (Chin). Phagemids were plated and picked at the Clemson  
 University Genomics Institute (CUGI) (Begum, Palmer,  
 Frisch, Atkins and Wing). Plasmid DNA preparations, DNA  
 sequencing and sequence analysis were performed at CUGI  
 (Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The  
 sequence has been trimmed to remove vector sequence and

contains a minimum of 100 bases of phred value 20 or  
 above. For more details on library preparation and  
 sequence analysis see  
<http://www.genome.clemson.edu/projects/barley>. To order  
 this clone see <http://www.genome.clemson.edu/orders> Also  
 see Close TJ, Wing R, Kleinbols A, Wise R (2001)  
 Genetically and physically anchored EST resources for  
 barley genomes. Barley Genetics Newsletter 31:29-30.  
 (<http://wheat.pw.usda.gov/99pages/Bgn/31/clover.html>)  
 BASE COUNT 162 a 135 c 185 g 143 t 2 others  
 ORIGIN  
 Query Match 24.5%; Score 417; DB 13; Length 627;  
 Best Local Similarity 84.5%; Pred. No. 6.2e-97;  
 Matches 479; Conservative 0; Mismatches 87; Indels 1; Gaps 1;  
 QY 888 GCATATATTAACAGTGTATCAC GAGGAAAGGACAGACATTATTTCTTCAAGCATTTTA 946  
 Db 18 GCTATATATTAACAGTGTATCACGAGGAGGAAAGACAGACATTATTTCTTCAAGCATTTCA 77  
 QY 947 TCAGGCTTGACGCTCATCCACACGAGAACTTAAGTCTAGATATACGCTGTAGT 1006  
 Db 78 TCAGGCTTGACGCTCATCCAAATCAGAGTATGAGTCCCAAGTGCATGCTGATG 137  
 QY 1007 TGTGGAGTGTATGTATGCTGTACACCAATTGGAATGATGATGCTGATGCTTGTGCT 1066  
 Db 138 TGTGGAGTGTATGATGATGCTGTACCAAGGAGGAGACACAGATTACGTGAATTTGTGCG 197  
 QY 1067 GAAGAACAGATTCATGACCGTGTCCATTTTGTGAACAAACATTTGGAGAGGCCCTTA 1126  
 Db 198 GAAGAAATGGAATCATGATGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 257  
 QY 1127 CTGGGACGATATGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1186  
 Db 258 TTGGGCTGGAATGATGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 317  
 QY 1187 GATTAACATTTGAAGCAATGCAATTCATTCAGTGTGAGTGTGAGGACGCTGTGAGGAGAC 1246  
 Db 318 GATTAACATTTGAAGCAATGCAATTCATTCAGTGTGAGTGTGAGGAGGAGGAC 377  
 QY 1247 CAGGAGATGCTGCTGAGAGGCTGCACTGAGGCTTCTGATCTGCTGAGGAGAGGAGGCT 1306  
 Db 378 CACGGAATGCTGCTGAGAGGCTGCACTGAGGCTTCTGATCTGCTGAGGAGAGGAGGCT 437  
 QY 1307 GCGGCTCTTGGCAAGAACATGCTGACCTGCAAGCCAGCCAGAGAGGCTTCAT 1366  
 Db 438 CAGGCTCTTGGCAAGAACATGCTGACCTGCAAGCCAGCCAGAGAGGCTTCAT 497  
 QY 1367 GGGGAAAGGCTATGAGAGGCTGAGGAAATGTTATGAGACCAATGCTGTAGAG 1426  
 Db 498 GGGAAACAGGCTATGAGAGGCTGAGGAAATGTTATGAGACCAATGCTGTAGAG 557  
 QY 1427 GATGCGGCGGTGTGAAGATGTCT 1453  
 Db 558 CATGCGCAGGCTGTGAAGAGATGTCT 584  
 RESULT 10  
 BGS60599 439 bp mRNA linear EST 10-APR-2001  
 LOCUS RH122.59.D08.b1.A003 Rhizome2 (RH122) Sorghum propinquum cDNA, mRNA  
 DEFINITION sequence.  
 ACCESSION BGS60599  
 VERSION BGS60599.1 GI:13589597  
 KEYWORDS EST.  
 SOURCE Sorghum propinquum.  
 ORGANISM Sorghum propinquum.  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC  
 clade; Panicoideae; Andropogoneae; Sorghum.  
 1 (bases 1 to 439)  
 Cordomer-Pratt, M.-M., Gingle, A., Paterson, A., Sudman, M. and Pratt,  
 L.H.

TITLE An EST database from Sorghum: Sorghum propinquum rhizomes  
JOURNAL  
COMMENT Unpublished (2000)  
Contact: Cordemier-Pratt MM  
Laboratory for Genomics and Bioinformatics  
The University of Georgia, Department of Plant Biology  
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA  
Tel: 706 542 1860  
Fax: 706 583 0210  
Email: mmpratt@uga.edu  
Seq primer: JEN REV  
High quality sequence stop: 429  
POLYA=No.

FEATURES  
source  
1. .439  
/organism="Sorghum propinquum"  
/db\_xref="taxon:132711"  
/clone\_lib="Rhizome2 (RHIZ2)"  
/note="Organ: Rhizomes; Vector: pBluescript II from Lambda  
Zap II; Site 1: XhoI; Site 2: EcoRI; The library was made  
from poly-A RNA in the cloning vector Lambda Zap II.  
Clones to be sequenced were prepared by mass excision."  
Location/Qualifiers

BASE COUNT 100 a 107 c 123 g 109 t  
ORIGIN

Query Match 23.6%; Score 401.6; DB 12; Length 439;  
Best Local Similarity 95.6%; Pred. No. 5.4e-93;  
Matches 413; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 206 GCGAGACCTTGGCCAGCGCGCTGGGAGACCGCGCCCTTACACCGCGCTGGCAGCGG 265  
DB 8 GCGAGACCTTGGCCAGCGCGCTGGGAGACCGCGCCCTTACACCGCGCTGGCAGCGG 67  
QY 266 GAGTCCCTCGGCTTCATGAGTCCAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 325  
DB 68 GAGTCCCTCGGCTTCATGAGTCCAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 127  
QY 326 CTCTGGTGGTCACTTTTACTGATGGAATGACATTTCTTGGAGCATGTTGGCTTGGCA 385  
DB 128 TTGGGGTGGCCACTTTTACTGATGGAATGACATTTCTTGGAGCATGTTGGCTTGGCA 187  
QY 386 AGTGTGTGTGATTAACAACACAGATCAACAAGAAACAAATGATGATACATATAGCTTGA 445  
DB 188 AGTGTGTGTGATTAACAACACAGATCAACAAGAAACAAATGATGATACATATAGCTTGA 247  
QY 446 GCAATGAGATTTGAACATGAGAGTGGTTCACGCTAGAGACAGAGGAGATGTA 505  
DB 248 GCAATGAGATTTGAACATGAGAGTGGTTCACGCTAGAGACAGAGGAGATGTA 307  
QY 506 TATTTGCTTAAAGCTGATCTGTTATCTTAAACATGCTGTTGCTGGCAAGTGGCTTGA 565  
DB 308 TATTTGCTTAAAGCTGATCTGTTATCTTAAACATGCTGTTGCTGGCAAGTGGCTTGA 367  
QY 566 CCGTGTCTGAAGAATGATGTTCTTAAAGTCTTCCGAATTTTGTGTGATGATCCATA 625  
DB 368 TCGTGTCTGAAGAATGATGTTCTTAAAGTCTTCCGAATTTTGTGTGATGATCCATA 427  
QY 626 AATGCTGGGCA 637  
DB 428 AATGCTGGGCA 439

RESULT 11  
B0148371  
LOCUS B0148371 685 bp mRNA linear EST 24-APR-2002  
DEFINITION NF067F07FL1063 Developing flower Medicago truncatula cDNA clone  
B0148371  
B0148371  
B0148371.1 GI:20285430  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.  
1 (bases 1 to 685)  
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,  
Flores, H.R., Iman, J.T., Weller, J.W. and May, G.D.  
Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
Medicago truncatula flower library  
Unpublished (2001)  
Contact: May GD  
Plant Biology Division  
The Samuel Roberts Noble Foundation  
2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
Tel: 580 221 7391  
Fax: 580 221 7380  
Email: gdmay@noble.org  
Insert Length: 685 Std Error: 0.00  
Plate: 067 row: F column: 07  
Seq primer: TCACACAGGAACGCTATGAC.  
Location/Qualifiers

FEATURES  
source  
1. .685  
/organism="Medicago truncatula"  
/db\_xref="taxon:3880"  
/clone\_lib="NF067F07FL"  
/clone\_lib="Developing flower"  
/tissue\_type="Developing flowers"  
/dev\_stage="developmentally pooled"  
/note="Vector: Lambda Zap; cDNA was prepared from polyA+  
enriched, pooled samples of equivalent amounts of total  
RNA from very young, developing, fully-opened flowers and  
flowers transitioning into pods. The cDNA was  
directionally ligated into the Uni-Zap XR vector  
(Stratagene) and packaged using the GigaPack III Gold  
packaging extracts. Phagemids containing cDNA inserts were  
in vivo excised from the recombinant Uni-Zap XR vector  
using Exsist helper phage and the E. coli strain  
XL1-Blue MRF' (Stratagene). Excised plasmids were plated  
using SOLR cells."

BASE COUNT 202 a 115 c 162 g 203 t 3 others  
ORIGIN

Query Match 23.2%; Score 395; DB 14; Length 685;  
Best Local Similarity 74.8%; Pred. No. 3.2e-91;  
Matches 505; Conservative 0; Mismatches 168; Indels 1; Gaps 1;

QY 510 GCTTAAAGCTGATGCTGTTATCTTAAACATGCTGTTGCTGGCAAGTGGCTT-GACCC 568  
DB 12 GCTTAAAGCTGATGCTGTTATCTTAAACATGCTGTTGCTGGCAAGTGGCTGATGC 71  
QY 569 TGTTCGAAAGATCAAGTTCCTTAAAGCTTCGGAAGATTTGTGGTGGATCCATGAAT 628  
DB 72 GCTTAAAGCTGATGCTGTTATCTTAAACATGCTGTTGCTGGCAAGTGGCTGATGC 131  
QY 629 GGTGGGCAATTAAGTGAATATGCAACATCTTCCCTTGTGCTGGAGCCAT 688  
DB 132 GCGAGGCAATTAAGTGAATATGCAACATCTTCCCTTGTGCTGGAGCCAT 191  
QY 689 GATTGATTCATACAGCGGCTGAGTATGGAATAGCAGGACTGAGGAGCTGGAAT 748  
DB 192 GATTGATTCATACAGCGGCTGAGTATGGAATAGCAGGACTGAGGAGCTGGAAT 251  
QY 749 ACGATGCCCAAACTTATGTTGTCACCTGGGGAATAGTAAGAACTAATGAGATGGC 808  
DB 252 TAAATGCTGGAAGAACTATGTTGATCATCTTGAATATGCAAGAACTAATGAGATGGC 311  
QY 809 TGAAGCAATGTCGCAAGAGAGTCTAGGGAACATTTGTGATACCTCGTGGAGTACG 868  
DB 312 AGAAGATAGTGAAGAGAGGTTCTTCGTGAGCATGTTGGAATCTCTTGGAGTAG 371  
QY 869 GAGTGGAGTCTCTGTTGCAATATTAACAGTGTATCAGAGAAAGGCAAGACTT 928  
DB 372 AATGATGATCTATCTTTGTCATCAATTAATGATGTTTACAGGTGTAAGGCAAGACTT 431

Qy 929 ATTCTTCAAGCATTTTATCAGGCTTTGGAGCTCATCCACAGCAAACTTAAGTCC 988  
 |||||  
 Db 432 ATTCCTTATGCTCTTTCATTAAGATTGGACGCTCATTCAGAGAAAGTCAAGTCC 491  
 |||||  
 Qy 989 TAGAATACATGCTGTATGTTGGAGATGATTAATGCTACACCAATTTGAATCTCA 1048  
 |||||  
 Db 492 ATCTTTCATGCTATGATGTTAGAGAGATGATGATGCTACACCAAAAGTTGAATGSA 551  
 |||||  
 Qy 1049 GTTACGCTGCTTGTGTGAGAGAACAGATTGATGACGCTGCTCATTTTGTGACAAAGAC 1108  
 |||||  
 Db 552 ACTCGCAATTTGTTTACAGAGAAAAGATTGACGACGCTGTATTTGTTAAGAAAC 611  
 |||||  
 Qy 1109 ATTGGCAGTGGCCCCCTTACTTGGCAGCAATTGATGCTGTTTCAAGATTTCAGGCGG 1168  
 |||||  
 Db 612 CCTGACAGTGGCTCTTACCTTACCTTATGATGTTCTTGTTCANAAATTCAGANGACN 671  
 |||||  
 Qy 1169 TGGAGAAATGCTTG 1182  
 |||||  
 Db 672 GGGAGAAATGTTGG 685  
 |||||

RESULT 12  
 BI271361 641 bp mRNA linear EST 18-JUL-2001  
 LOCUS NF050A09FL1069 Developing flower Medicago truncatula cDNA clone  
 DEFINITION NF050A09FL 5', mRNA sequence.  
 ACCESSION BI271361  
 VERSION BI271361.1 GI:14879765  
 KEYWORDS EST.  
 SOURCE barrel medic.  
 ORGANISM Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;  
 Medicago.  
 1 (bases 1 to 641)

REFERENCE  
 AUTHORS Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,  
 Flores, H.R., Imman, J.T., Weller, J.W., and May, G.D.  
 TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation  
 Medicago truncatula flower library  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: May GD  
 Plant Biology Division  
 The Samuel Roberts Noble Foundation  
 2510 Sam Noble Parkway, Ardmore, OK 73402, USA  
 Tel: 580 221 7391  
 Fax: 580 221 7380  
 Email: gdmay@noble.org  
 Insert Length: 641 Std Error: 0.00  
 Plate: 050 row: A column: 09  
 Seq primer: TCACACAGCAAGACGCTATGAC.  
 Location/Qualifiers  
 1..641  
 /organism="Medicago truncatula"  
 /db\_xref="taxon:3880"  
 /clone="NF050A09FL"  
 /clone\_lib="Developing flower"  
 /tissue\_type="Developing flowers"  
 /dev\_stage="Developmentally pooled. Contains a mixture of  
 very young, developing, fully-opened flowers and flowers  
 in early transition into pods."  
 /note="Vector: Lambda Zap; cDNA was prepared from polyA+  
 enriched, pooled samples of equivalent amounts of total  
 RNA from very young, developing, fully-opened flowers and  
 flowers transitioning into pods. The cDNA was  
 directionally ligated into the Uni-Zap XR vector  
 (Stratagene) and packaged using the Gigapack III Gold  
 packaging extracts. Phagemids containing cDNA inserts were  
 in vivo excised from the recombinant Uni-Zap XR vector  
 using ExSist helper phage and the E. coli strain  
 XL1-Blue MRF' (Stratagene). Excised plasmids were plated  
 using SOLAR cells."

## FEATURES

source

1..641  
 /organism="Medicago truncatula"  
 /db\_xref="taxon:3880"  
 /clone="NF050A09FL"  
 /clone\_lib="Developing flower"  
 /tissue\_type="Developing flowers"  
 /dev\_stage="Developmentally pooled. Contains a mixture of  
 very young, developing, fully-opened flowers and flowers  
 in early transition into pods."  
 /note="Vector: Lambda Zap; cDNA was prepared from polyA+  
 enriched, pooled samples of equivalent amounts of total  
 RNA from very young, developing, fully-opened flowers and  
 flowers transitioning into pods. The cDNA was  
 directionally ligated into the Uni-Zap XR vector  
 (Stratagene) and packaged using the Gigapack III Gold  
 packaging extracts. Phagemids containing cDNA inserts were  
 in vivo excised from the recombinant Uni-Zap XR vector  
 using ExSist helper phage and the E. coli strain  
 XL1-Blue MRF' (Stratagene). Excised plasmids were plated  
 using SOLAR cells."

BASE COUNT 191 a 105 c 151 g 191 t 3 others  
 ORIGIN

Query Match 22.0%; Score 375.2; DB 13; Length 641;  
 Best Local Similarity 75.8%; Pred. No. 4,1e-86;  
 Matches 475; Conservative 0; Mismatches 151; Indels 1; Gaps 1;

Qy 602 GAAGATTTTGTGTGATGATCAAGAAATGCGTGATTTAGGTGAATATGCA 661  
 |||||  
 Db 1 GAAGTTTATGATGATGATCAAGAAATGCGTGATTTAGGTGAATATGCA 60  
 |||||  
 Qy 662 AACTTCTCCCTTTGTGTGAGGCAATGATTTTCATCAAGAGGCTGAGTA-TTGA 720  
 |||||  
 Db 61 GCACCTCCCTTTGTGTGAGGCTGATGATTTTCATCAAGAGGCTGAGTA 120  
 |||||  
 Qy 721 ATACAGAGCTAGCGATGCGCTGAAAATACAGATGCCACAACTTATGTTGTA 780  
 |||||  
 Db 121 AGAATAGACCAAGAGAGATTAGGAATTAATGCTGAACTTATGTTGATCTTG 180  
 |||||  
 Qy 781 GGAATAGTAAGAACTATGAGAGTTGTTGAAAGACAAATGTCGAGAGAGCTTAC 840  
 |||||  
 Db 181 GAATAGCAAGAGAACTTATGAGAGTTGCGAAGATAGTATGCAAGAGGTTCTG 240  
 |||||  
 Qy 841 AACATATGCTGAATCCCTTGAAGTACGAGTGAAGATCTCTCTTTGCAATATA 900  
 |||||  
 Db 241 AGCATGTAGGGAATCTCTTGAAGTGAAGATGATCTTATTTGCAATATA 300  
 |||||  
 Qy 901 GTGATACAGAGAAAGGAGCAAGACTTATCTTCAAGCAATTTTACGGCTTGAG 960  
 |||||  
 Db 301 GTGTTTACAGTGTGTAAGGCAAGATCTATCTCTTATGCTTCAATATAAG 360  
 |||||  
 Qy 961 TCATCCACAGAGAACTTAAAGTCCCTAGAAATCATCTGTAAGTTGGAGAGTATG 1020  
 |||||  
 Db 361 TCATTCAGAGAGAAAGTCAAGTCCCATCTTGCATATGATGTTGAGAGATATA 420  
 |||||  
 Qy 1021 TTAATGCTAGACCAATTTGAGACTCACTTACGTAAGTTGTGTAAGAAACGATTC 1080  
 |||||  
 Db 421 TGAATGCTACCAAAAGTTGAAGTGAAGTCCGCAATTTGTTACAGAGAAAGATTC 480  
 |||||  
 Qy 1081 ATGACGCTGCTCATTTTGTGAACAAGACTTGGCAGTGGCCCTTACTTGGAGCAATTG 1140  
 |||||  
 Db 481 AGACGCGTGTCAATTTGTGAACAAACCCCTGCGAGTGGCTCTTCAAGGTTCTATG 540  
 |||||  
 Qy 1141 ATGCTTGTGAGATTCAGAGGCGGTGAGATGCTTTGGAAGATTAATTAATTGAAG 1200  
 |||||  
 Db 541 ATGTTCTGTGTAAGATTCAGAGGAGONGGAGATGTTTGAAGATTAATTAAG 600  
 |||||  
 Qy 1201 CAATGCAATTCAGATTGCCAGTATTGG 1227  
 |||||  
 Db 601 CANTGGCAATTCGCTGCTGTACTGG 627  
 |||||

RESULT 13  
 AM186747 643 bp mRNA linear EST 30-NOV-1999  
 LOCUS BNGH113725 Six-day Cotton fiber Gossypium hirsutum cDNA 5', mRNA  
 DEFINITION BNGH113725 Six-day Cotton fiber Gossypium hirsutum cDNA 5', mRNA  
 sequence.  
 ACCESSION AM186747  
 VERSION AM186747.1 GI:6461183  
 KEYWORDS EST.  
 SOURCE upland cotton.  
 ORGANISM Gossypium hirsutum  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.  
 1 (bases 1 to 643)

REFERENCE  
 AUTHORS Blewitt, M., Matz, E.C., and Burr, B.  
 TITLE ESTs from developing cotton fiber (1999b)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Ben Burr  
 Biology Department  
 Brokhamen National Laboratory  
 Upton, NY 11973, USA

Tel: 516-344-3396  
 Fax: 516-344-3407  
 Email: burr@bnl1.bnl.gov  
 Seq primer: T3 Primer.  
 Location/Qualifiers  
 1..643  
 /organism="Gossypium hirsutum"  
 /cultivar="Acala Maxxa"  
 /db\_xref="taxon:3635"  
 /clone\_lib="Six-day Cotton fiber"  
 /issue\_type="mature fiber"  
 /dev\_stage="six days post anthesis"  
 /lab\_host="XLI-Blue"  
 /note="Vector: Bluescript II KS+"

BASE COUNT 193 a 112 c 159 g 179 t  
 ORIGIN

Query Match 21.2%; Score 360.4; DB 10; Length 643;  
 Best Local Similarity 73.0%; Pred. No. 2,8e-82;  
 Matches 463; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 430 TCACATATGCTTGAGCATAGATGTCACATGAGTGACGTTTACCGACTAGAG 489  
 Db 2 TCATTTACAGTTTAAACATAGATGTGATCGAGGGTTCAGTTTCTCTCCAAAG 61  
 QY 490 GACAGAGGCACTGATATTGCTCTAAAGTGATCTGTTATCTTAAACACTGCTG 549  
 Db 62 GTCAAGAGCTATATATCTGCTTAAAGCTGATTTGTAAGTTGAACCTCGGTTG 121  
 QY 550 CTGGCAAGTGCTTACCCCTGTTCTGAAGATCATGTTCTTAAAGCTCTCCGAAGTTT 609  
 Db 122 CTGAAAGGCTGAGTGCTGTTCTTAAAGACATGTTGATGTTCTTCAAGGTTG 181  
 QY 610 TGTGTGATGCCAATGATGCTGAGCATTTAGTTGAATATGTCACATCTTC 669  
 Db 182 TCTGTGATTCATGAATGAGGACATTTACTCAAGCTAGAAATGTTAAGACCTCC 241  
 QY 670 CTTTGTGCTGAGCCCATGATGATTCATACAGGCTGATTTGAATAGCAGGA 729  
 Db 242 CTTTGTGAGTGCCATGATGATTCACACACAGCAATTTCTGGAAGAACCGGA 301  
 QY 730 CTAGCATGCTCGTGAATACAGATGCCAACAATTTATGTTTCACTGGGGAATAGTA 789  
 Db 302 CTGGGAAAGATTTAGGATCAAAATGCTGAAACATGTTTGACATTAAGAAACAGA 361  
 QY 790 AAGACTATGGAATGCTGAGAGCAATGTCGAAAGAGTCCACGGAACATATTC 849  
 Db 362 AGAGCTATGGAATGCTGAGAAATGATGTCGAAAGGTTCTCTGAGCATGTTG 421  
 QY 850 GTGAATCCTTGGAGTACGAGTGAAGATCTCTGTTGCAATATATTAACAGTATAC 909  
 Db 422 GTGAATCTCTGGCTGAAAAATGAAGATCTGTTGCTTTATTAATAGTATTC 481  
 QY 910 GAGGAAAGGACAAGACTTATTTCTTCAAGCTTTTATCAGGCTTTGACGCTCAAC 969  
 Db 482 GAGGAAAGGTCAGACTATTTCTCTGCTTCTTCAAGAAATTTGAAGTTGATCAG 541  
 QY 970 ACGAAGACTTAAATGCTTGAATATCATGCTGATGTTGGAGAGTATGTTATGCTC 1029  
 Db 542 AGGAAAGCTAAAGGTCGACCAATTCATGCGCGATAGTGGAGAGTCAATGATGCTC 601  
 QY 1030 AGACCAATTTGAGACTCACTTACGTAAGTTTGT 1063  
 Db 602 AAACAAGTTTGAACAGAGCTGCTTACTTGT 635

RESULT 14  
 AM257867 375 bp mRNA linear EST 30-MAR-2000  
 LOCUS 687063C09.y1 687 - Early embryo from Delaware Zea mays cDNA, mRNA  
 DEFINITION  
 sequence.  
 ACCESSION AM257867  
 VERSION AM257867.1 GI:6626335

KEYWORDS EST.  
 Zea mays.  
 SOURCE Zea mays.  
 ORGANISM Zea mays.

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PAC  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 375)  
 Walbot, V.  
 Maize ESTs from various cDNA libraries sequenced at Stanford  
 University

JOURNAL Unpublished (1999)  
 CONTACT Walbot, V.  
 DEPARTMENT Department of Biological Sciences  
 INSTITUTION Stanford University  
 ADDRESS 855 California Ave, Palo Alto, CA 94304, USA  
 TEL: 650 723 2227  
 FAX: 650 725 8221  
 EMAIL: walbot@stanford.edu  
 PLACE: 687063 row: C column: 09.

FEATURES Location/Qualifiers

SOURCE

1..375  
 /organism="Zea mays"  
 /cultivar="Illinois High Oil"  
 /db\_xref="taxon:4577"  
 /clone\_lib="687 - Early embryo from Delaware"  
 /issue\_type="embryo"  
 /dev\_stage="14, 21, 28, and 35 days after pollination"  
 /lab\_host="E. coli SOLR"  
 /note="Organ: embryo; Vector: pBluescript SK, Site 1: XhoI  
 ; Site 2: EcoRI; Library was prepared by Scatagene using  
 the Uni-ZAP XR system (Stratagene BN37328-12). Clones  
 were picked by a Q-bot after blue/white selection  
 (ampicillin resistance - use 100 micrograms/microliter).  
 Developed from a pool of equal amounts of RNA from  
 developing embryos sampled at 14, 21, 28 and 35 days after  
 pollination of the Illinois High Oil Maize strain Cycle  
 90. This closed strain has been selected for high oil  
 concentration for 90 generations and originates from the  
 1890s era open pollinated variety Burr's White"  
 80 a 104 c 111 g 77 t 3 others

ORIGIN

Query Match 20.1%; Score 342.4; DB 10; Length 375;  
 Best Local Similarity 96.0%; Pred. No. 1.1e-77;  
 Matches 360; Conservative 0; Mismatches 14; Indels 1; Gaps 1;

QY 155 GGCCTGCGACCTCCGATCCACAGCAGAGGTTCTCTCCGCGTCCCTGCGAGACCC 214  
 Db 1 GGCCTGCGACCTCCGATCCACAGCAGAGGTTCTCTCCGCGGAGCTCTGGAACCC 60  
 QY 215 TT-GCAGCGGCGTGGGGAACCCGCGCCTCAACACCGCGTGCAGCGGAGTCCC 273  
 Db 61 TTGGGAGCGGCGTGGGGAACCCGCGCCTCAACACCGGTCGCAAGCGGAGTCCC 120  
 QY 274 TGGGTTTATGAGTCCAGCTCGTCTCTCTCCCAAGACTCTCCCTCTGCTG 333  
 Db 121 TGGGTTTATGAGTCCAGCTCGTCTCTCTCCCAAGACTCTCTCTCTGCTG 180  
 QY 334 GTCCACTTTTATGATGATTAAGATTTCTTCTGAGGAGTGTGCTGCAAGTGTGT 393  
 Db 181 GTCCACTTTTATGATGATTAAGATTTCTTCTGAGGAGTGTGCTGCAAGTGTGT 240  
 QY 394 GATTAACCAACCAAGATATCAAGAAACAAATGATGATCTTATGTTGAGCATAGA 453  
 Db 241 GATTAACCAACCAAGATATCAAGAAACAAATGATGATCTTATGTTGAGCATAGA 300  
 QY 454 TGTGAACATGAGTGCAGGTTTACAGCTAGAGGACAGAGGAGTATTTGCTC 513  
 Db 301 TGTGAACATGAGTGCAGGTTTACAGCTAGAGGACAGAGGAGTATTTGCTC 360  
 QY 514 TAAAGCTGATCTGG 528  
 Db 361 TAAAGCTGATCTGG 375

```
RESULT 15
B0464581
LOCUS B0464581 437 bp mRNA linear EST 30-MAY-2002
DEFINITION Hf Hordeum vulgare cDNA clone HF02L01 3-PRIME, mRNA
sequence.
ACCESSION B0464581
VERSION B0464581.1 GI:21272363
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
1 (bases 1 to 437)
Radchuk,V., Zhang,H., Weschke,W., Potokina,E. and Wobus,U.
Barley ESTs from developing seeds
Unpublished (2002)
Contact: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 437 Std Error: 0.00
Plate: 2 row: 1 column: 1
Seq primer: M3rev.
FEATURES
source
1..437
location/Qualifiers
/organism="Hordeum vulgare"
/cultivar="Barke"
/db_xref="taxon:4513"
/clone="HF02L01"
/clone_1b="HF"
/cissue_type="caryopsis"
/dev_stage="developing caryopsis, 16-25 DAF (days after
flowering)"
/lab_host="XL10-Gold"
/note="Vector: Bluescript SK+, Site 1: EcoRI (5'-end of
cDNA); Site 2: XhoI (3'-end of cDNA); developing caryopsis
, 16-25 DAF(days after flowering) Due to a cloning artefact
caused by the kit, in most cases the EcoRI site is NOT
present, as well as the EcoRI adapter used for cloning. To
excise the insert, restriction sites upstream EcoRI should
be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the
cloning system used Blue/white selection for recombinats
is not 100% reliable.Average insert size is 940 bp"
BASE COUNT 109 a 95 c 124 g 109 t
ORIGIN
Query Match 19.2%; Score 327.2; DB 14; Length 437;
Best Local Similarity 84.4%; Pred.No.9.7e-74;
Matches 368; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
QY 917 GGGACAAGCTATTCTTCAGCATTTTACAGCTTTGACCTCATCAACGAGAA 976
DB 2 GGCACAGGCTTATTCTTCAGCATTTTACAGCTTCACAACTCAAAAATCAGAA 61
QY 977 ACTTAAGTGCCTAGATATCATGCTGTAGTTGTGGAAGTATGTTAATGCTCAGACCA 1036
DB 62 GTTAGAAGTGCCTAAGTGCATGCTGTAGTTGTGGAAGTATGATATGCTCAGACAA 121
QY 1037 ATTGAGACTGAGTAGTACTTGTGTGTAAGAACAGATTCATGACCGTGCCATT 1096
DB 122 GTTTGAGACACAGTTACGTGATTTGTGGCAGAGATGGATCCATGACCGTGTCATT 181
QY 1097 TGTGAACAAGACATTTGAGTGCCTTACTTGTGCAAGCAATGATGCTTGTTCAGA 1156
DB 182 TGTCAACAAGACGTTGGAGTGCACCTTATTGGCTGCAATGATGCTTGTTCAGA 241
QY 1157 TTCTCAGGGCCGTGAGATGCTTGTGAAGATAAACAATTGAAGCATGCAATGATT 1216
```

```
DB 242 CTCTCAGCCCGTGGAGATGCTTTGGAAGATTAACATTTGAAGCATGGCTTCAAGTT 301
QY 1217 GCCAGTATTGGGACAGCGCTGCTGGAAGGACACGAGATGTCCTGACGGCTGACTGG 1276
DB 302 GCCAGTGTGGGTACGGCTGCTGGAAGGACACGAAATGCTCTTGGATGGTTCAACCG 361
QY 1277 CTTTGTGATCCTGCTGGAAGGAGGCGCTGCGCTCTTGCAGAAAGAACATGTCAGACT 1336
DB 362 CTTTGTGACCGCTGCTGGAAGGAGGCGCTCAGCGCTCTGCGCAGGAATGTTAAGCT 421
QY 1337 CGCAAGCCAGCGCCAG 1352
DB 422 TGGAGCCATGTTGAG 437
```

Search completed: July 4, 2003, 22:05:50  
Job time : 1711 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: June 30, 2003, 16:18:26 ; Search time 72 seconds  
(without alignments)  
875.383 Million cell updates/sec

Title: US-09-938-294-45

Perfect score: 2297  
Sequence: 1 MAKTFEFAVAAGRGQPVH.....ERIAVLKDYLRKQSHSR 473

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database :

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1563.5	65.2	457	21	AA028811 Arabidopsis thalia
2	1563.5	65.2	463	21	AA028810 Arabidopsis thalia
3	1559.5	65.1	449	21	AA028812 Arabidopsis thalia
4	487	20.3	697	23	AB092331 Hericidially activ
5	482.5	20.1	670	23	AB091330 Hericidially activ
6	482.5	20.1	670	23	AB091331 Hericidially activ
7	179.5	7.5	424	22	AB057933 Drosophila melanog
8	161.5	6.7	385	21	AAV68959 Osg2 protein whic
9	161	6.7	405	21	AA043119 Arabidopsis thalia
10	157.5	6.6	397	21	AA043120 Arabidopsis thalia

11	157	6.5	403	21	AA024139 Arabidopsis thalia
12	153.5	6.4	395	21	AA024140 Arabidopsis thalia
13	152.5	6.4	368	21	AA092331 An Escherichia coli
14	152.5	6.4	368	22	AB052953 Escherichia coli p
15	152.5	6.4	368	22	AB052987 Escherichia coli p
16	150	6.3	387	22	AA090136 C glutamicum prote
17	135	5.6	402	22	AA096378 Putative P. abyssi
18	134.5	5.6	416	22	AA027831 Human full-length
19	134.5	5.6	416	22	AA093159 Human polypeptide
20	134.5	5.6	416	22	AA094024 Human protein sequ
21	134.5	5.6	416	23	AB080682 Human transferase
22	134	5.6	258	21	AA024141 Arabidopsis thalia
23	133.5	5.6	390	22	AA076597 Corynebacterium gl
24	133.5	5.6	409	22	AA090984 C glutamicum prote
25	133.5	5.6	416	22	AA093589 Human polypeptide
26	133	5.5	260	21	AA043121 Arabidopsis thalia
27	130.5	5.4	371	22	AA096338 Putative P. abyssi
28	129.5	5.4	335	22	AA096478 C glutamicum prote
29	128.5	5.4	418	22	AA090197 Corynebacterium gl
30	127	5.3	381	22	AA092142 Human novel glycos
31	127	5.3	381	22	AA076598 S. thermophilus exo
32	126.5	5.3	416	22	AA090927 S. thermophilus exo
33	125.5	5.2	372	18	AA022175 Enzyme Eps2 involv
34	125.5	5.2	384	18	AA014076 Enzyme Eps2 involv
35	125	5.2	371	21	AA043076 Streptococcus pneu
36	125	5.2	371	21	AA043078 Streptococcus pneu
37	124.5	5.2	350	19	AA062736 Streptococcus pneu
38	123.5	5.2	367	21	AA081719 Streptococcus pneu
39	119	5.0	387	23	AA083361 Novel human enzyme
40	117	4.9	285	22	AA023781 Human polypeptide
41	114	4.8	285	23	AA089378 Amino acid sequenc
42	113	4.7	720	21	AA069235 Galactosyl transfe
43	112.5	4.7	397	20	AA080300 Human PFO1298 (UHQ
44	111	4.6	323	21	AA093977 Human PRO polypept
45	111	4.6	323	22	AA029171

## ALIGNMENTS

RESULT 1  
AA028811 standard; Protein: 457 AA.

ID	AA028811	standard; Protein: 457 AA.
XX	AC	AA028811;
XX	AC	17-OCT-2000 (first entry)
XX	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 34169.
XX	DE	Protein identification; signal transduction pathway; metabolic pathway;
XX	KW	hybridisation assay; genetic mapping; gene expression control; promoter;
XX	KW	termination sequence.
XX	OS	Arabidopsis thaliana.
XX	PN	EP1033405-A2.
XX	PD	06-SEP-2000.
XX	PF	25-FEB-2000; 2000EP-0301439.
XX	PR	25-FEB-1999; 99US-0121825.
XX	PR	05-MAR-1999; 99US-0123180.
XX	PR	09-MAR-1999; 99US-0123548.
XX	PR	22-MAR-1999; 99US-0125788.
XX	PR	25-MAR-1999; 99US-0126264.
XX	PR	29-MAR-1999; 99US-0126785.
XX	PR	01-APR-1999; 99US-0127462.
XX	PR	06-APR-1999; 99US-0128234.
XX	PR	08-APR-1999; 99US-0128714.
XX	PR	16-APR-1999; 99US-0129845.
XX	PR	19-APR-1999; 99US-0130077.



PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161355.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 65.2%; Score 1563.5; DB 21; Length 457;  
Best Local Similarity 69.0%; Pred. No. 1.8e-152;  
Matches 309; Conservative 63; Mismatches 65; Indels 11; Gaps 2;

QY 29 LTVVAASASTAGFLRLGALRDP-----DGRDPAALNTAVASGLPGLFPMRSKVL 80  
DB 8 LMTVLFLSVTCVILVSSFCSSISQFVEKKGSSA--AKROSNLDPFMRSKVL 64  
QY 81 LVSHSLSLSGCPILLMEIAFLLRHVSQVAVITNRSQETNDVTYSLEHRMLNHGVQLP 140  
DB 65 LVSHSLSLSXPLLMELIAFLLRGVADVAVITNOKPLEDEVVYSLEHMLDRGVQVIS 124  
QY 141 ARGEAVDIALKADVIINTAVAKMDPVLKDHVPKLPILMLWHEKRGHYFVKEVYK 200  
DB 125 AKGQKAVDTSLKADVLINTAVAGKMDAVLENVKKLPILMLWHEKRGHYFNADLVK 184  
QY 201 HLPVAGAMIDSHTTAETWNSRTSDRLTIQMPQTVVHLGNSKELMEVADNVARVIRE 260  
DB 185 HLPVAGAMIDSHATGKWKRRTOARIGIMPKTYVHLGNSKELMEVADNVARVIRE 244  
QY 261 HIRSLGVREDELLFATINSVSRGKQDPLQAFYQALQLOHEKLVPRHIAVAVVGS 320  
DB 245 HVRESLGVREDELLFATINSVSRGKQDPLQAFYQALQLOHEKLVPRHIAVAVVGS 304  
QY 321 NAQTKFETQLRDFVANKTIHDFVFNKTLAVAPYIAIDVIVONSQGRGECFRITTEA 380  
DB 305 SKQTKFETELRNFRREKLENFVFNKTLVAPYIAIDVIVONSQGRGECFRITTEA 364  
QY 381 NAFKLPVIGTAAAGTEIYLDGSGTGLHPAGKGVAPLAKNIYRLASHAFORSKMGEGY 440  
DB 365 NAFKLPVIGTAAAGTEIYLVNGTGLSHSKGKGVPLAKNIYRLAQVYLRIRKNGKGY 424  
QY 441 GGVKEMFMHHMARIAAVLKDVLRKSQ 468  
DB 425 ERVKEMFLEHMHSHRIASVLAKEVLOHAK 452

RESULT 2  
AAG28810  
ID AAG28810 standard; Protein; 463 AA.

XX AAG28810;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 34168.

XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.  
XX Arabidopsis thaliana.  
OS EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126254.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.



[illegible]

PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148585.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149375.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149920.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 30-AUG-1999; 99US-0151103.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158359.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 23-OCT-1999; 99US-0161404.  
PR 23-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 65.1%; Score 1559.5; DB 21; Length 449;  
Best Local Similarity 68.9%; Pred. No. 4.5e-152;  
Matches 308; Conservative 63; Mismatches 65; Indels 11; Gaps 2;

QY 30 LVVVAASSTAGFLLRGALRDP-----DGRGDAALNTRVAGSGPLGFMRSKIVL 81  
DB 1 WVLFLPSTVCTVILVRSFSCSISQFVEBKNGSSA---AKQSNPLDFMRSKIVL 57

QY VSHSLISGGPILLMELAFLLRHVGSQVWVITNORSQETDVYTSLEHRLNHGVOVLA 141  
DB VSHSLISXXPFLLMELFLRGGVADVWITNOKPLEDEBVSYSLEHMLDRGQVISA 117  
QY 142 RGEAVDIALKADVILNTAVAGKMLDPVLDKHPKVLPIKLWHEHNGHYFKVEYKH 201  
DB 118 KGQKAVDTSIKADLVLTAVAGKMDAVLKENVVKVLPKLWHEHNGHYFNADLVKH 177  
QY 202 LPFVAGMIDSHHTAEYNSRTPDLKIOMPTQYVHLGNSKEIMEVADNVARVIREH 261  
DB 178 LPFVAGMIDSHATAGYKNTQARLGIMPTVYVHLGNSKELMVAADSAAKVLREH 237  
QY 262 IRESLVGRSEDLIPAINSVSRKGQDLFLQAFYALOLIQHEKLVPIAHVAVVGSVN 321  
DB 238 VRESIGVRNDDLFGINSVSRGQDLFLRAFHESLERIKERKQVPTMHAVVGSIDMS 297  
QY 322 AQTETQLRDPFVKNQTHDRVFNKTLAVAPYLAIDLVLVONSQGRGECFGRITTEAM 381  
DB 298 KQTFETELRNPFREKLENFVHFVNKTLTVAPIAIDLVLVONSQARGECFGRITTEAM 357  
QY 382 AFKLPLVLTGTAAGTTEIYLDGSTGLHPAGKGVAPLAKNIVRLASHAQRVSGEKYXG 441  
DB 358 AFKLPLVLTGTAAGTTEIYVNGTGLHLSAGKEGVPILAKNIVKLATQVELRLRGKNGYE 417  
QY 442 RYKEMFMEHHMARIANAIVLKDVLRKSQ 468  
DB 418 RYKEMFLEHMSHRIASVLRKEVLQHAH 444

## RESULT 4

ABB92331 ID ABB92331 standard; Protein; 697 AA.

ABB92331;

31-MAY-2002 (first entry)

Herbicide(s) active polypeptide SEQ ID NO 1542.

Herbicide(s); plant; agriculture; herbicide.

Arabidopsis thaliana.

MO200210210-A2.

07-FEB-2002.

28-AUG-2001; 2001WO-EP09892.

28-AUG-2001; 2001WO-EP09892.

(FARB ) BAYER AG.

Tietjen K, Weidner M;

WPI; 2002-269010/31.

Identifying plant target proteins for herbicidally active compounds,  
comprising aligning and comparing nucleic acid or amino acid sequences  
from plant with nucleic acid or amino acid sequences from non-plant  
organisms -

Claim 5; SEQ ID NO 1542; 261pp + Sequence Listing; English.

The invention relates to identifying target proteins  
(ABB90790-ABB94016) for herbicidally active compounds, comprising  
aligning and comparing nucleic acid or amino acid sequences from plant  
with nucleic acid or amino acid sequences from non-plant organisms using  
suitable search parameters, where plant sequences having an E-value  
greater by a factor of 3 than the E-value of most similar non-plant  
sequences are selected. The polypeptides or nucleic acids encoding them  
are useful for identifying modulators. The identified modulators are

CC useful as herbicides.

XX Sequence 697 AA;

Query Match 20.3%; Score 487; DB 23; Length 697;  
Best Local Similarity 29.1%; Pred. No. 7.5e-41;  
Matches 147; Conservative 77; Mismatches 159; Indels 122; Gaps 13;

```

QY 52 CDGRDPAALNTAVASGSPFGFMR---SKLVLLVSHLSLSGGLPMLAEFLRHVSQ 108
DB 223 CDRKSD-----FKRLVMSRRFVLLFHELSMTGAPISWMEIASSELLSCAT 267
QY 109 VVMTNRSQETNDVYLSLHRMLNHGVQLPARGQEAVIDALKAADIYLNTAVAGKMD 168
DB 268 VYAVVLSRRG-----GLQELTRRIKVEDKGLSEKTPAMGADLVNGSVQASMTD 320
QY 169 PVLKDHVPKVLPKILMWIHEMRGHYFK---VEYVKHLPFVAGAMTDSHTTAAYWNS-R 222
DB 321 QYM-DHHPAGSQIAMWVWENRREYFDRAPKPVLDKVLILFL-----SEVQSKOMLTWC 373
QY 223 TSDRLKIOMPTVYVHLGNSKELMEVAEDVNA-----RRVLRHIRESLGV 268
DB 374 EBDHVKLR-SQPIVLPVSVNDELAFVAGVSSLNTPTLTQETMKEKKQKLRSEVTERGL 432
QY 269 RSEDLFPAIINSVSRGKQDLFLQAFYQALQ-----IQHEKLR 307
DB 433 TDKMLVMSLSINPGKQLLLESVALALERQTOEQVAKRNSKIINKLNGIRREKIS 492
QY 308 VPRHIAV-----VYGSQVN 321
DB 493 ISARRLRGSSRRMKITSPAVDNHPSVLSATGRKLLSGVNTOKDILLGVSQSKN 552
QY 322 AQTFEFTQLDRFVVKV-N-TIHDRVHFVNKTLANVAPYLAIDVLYONSQGRGECFGRTTEA 380
DB 553 -KVAVYEMKLSFLSNGNSVMTPTATTRVSLTSAADYVYVNSQGVETFGVTTIEA 611
QY 381 NAFKLPVLGTAAGTTEIYLDGSTGLHPAPKEGVAPLAKNIVPLASHAORVSGEKG 440
DB 612 MAYGLPVLGTDAGGTTEIYENAVTGLHPVGRAGNKVLAQNULPLNPSSTRLOQSQR 671
QY 441 GRVEMEMEHMAERIAVLAQDLVR 465
DB 672 EIVEKMTWKQMYKRFVDVLYKCMR 696

```

RESULT 5  
ABB91330  
ID ABB91330 standard; Protein, 670 AA.

```

XX AC ABB91330;
XX DT 31-MAY-2002 (first entry)
XX DE Herbicidally active polypeptide SEQ ID NO 541.
XX KM Herbicidal; plant; agriculture; herbicide.
XX OS Arabidopsis thaliana.
XX PN WO200210210-A2.
XX PD 07-FEB-2002.
XX PF 28-AUG-2001; 2001WO-EP09892.
XX PR 28-AUG-2001; 2001WO-EP09892.
XX PA (PARB ) BAYER AG.
XX PI Tietjen K, Weidler M;
XX DR WPI; 2002-269010/31.
XX

```

PT Identifying plant target proteins for herbicidally active compounds,  
PT comprising aligning and comparing nucleic acid or amino acid sequences  
PT from plant with nucleic acid or amino acid sequences from non-plant  
PT organisms -

PS Claim 5; SEQ ID NO 541; 261pp + Sequence listing; English.

XX The invention relates to identifying target proteins  
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising  
CC aligning and comparing nucleic acid or amino acid sequences from plant  
CC with nucleic acid or amino acid sequences from non-plant organisms using  
CC suitable search parameters, where plant sequences having an E-value  
CC greater by a factor of 3 than the E-value of most similar non-plant  
CC sequences are selected. The polypeptides or nucleic acids encoding them  
CC are useful for identifying modulators. The identified modulators are  
CC useful as herbicides.

XX Sequence 670 AA;

Query Match 20.1%; Score 482.5; DB 23; Length 670;  
Best Local Similarity 29.2%; Pred. No. 2.1e-40;  
Matches 148; Conservative 77; Mismatches 173; Indels 109; Gaps 13;

```

QY 38 STAGFLIR--GALRD-----PDGRDPAALNTAVASGSPFGFMR---SKLV 80
DB 193 ATYGKLLGFPFGLSDKVLWSPHRRSGTCDKSD-----FKRLVMSRRFV 237
QY 81 LVSHLSLSGGLPMLAEFLRHVSQVYVMTNRSQETNDVYLSLHRMLNHGVQLP 140
DB 238 LTFHELSMTGAPISWMEIASSELLSCATVSAVLSRRG-----GLQELSRRIKVE 290
QY 141 ARGQEAVIDALKAADIYLNTAVAGKMDPVLKDHVPKVLPKILMWIHEMRGHYFK----- 195
DB 291 DKGLSEKTPAMGADLVNGSVQASMTDQYMH-N-PAGGSQIAMWVWENRREYFDRAPK 349
QY 196 VEYVKHLPFVAGAMTDSHTTAAYWNS-RTSDRLKIOMPTVYVHLGNSKELMEVA----- 249
DB 350 LDRKYLIFL-----SEGSQRLMTWCEEHIKLR-SQPIVLPVSVNDELAFVAGIPSS 402
QY 250 -----EDNVARVLRHIRESLGVSEDLFPAIINSVSRGKQDLFLQAFYQALQ- 299
DB 403 LNTPTLSPEKMKVVKQILRESVRELGITDSMLVMSLSINPGKQLLLESIALALSE 462
QY 300 -----LQHEKLRVPRHIAV-----VYGS 318
DB 463 RQGESQNRNHKGIIRREKXVLSKHLRGSSRQMSVSLTDNGLRREKQELKVLGVSQS 522
QY 319 DVNAQTFEFTQLDRFVVKVNTIHDRVHFVNKTLANVAPYLAIDVLYONSQGRGECFGRTTE 378
DB 523 KSNRVGVYKEMLSFLSNGNSVMTPTATTRVSLTSAADYVYVNSQGVETFGVTTI 582
QY 379 EANAFLPVLGTAAGTTEIYLDGSTGLHPAPKEGVAPLAKNIVPLASHAORVSGEKG 438
DB 583 EANAFLPVLGTDAGGTTEIYENAVTGLHPSHSGRGNKELAHNULPLNPSSTRLOQSQR 642
QY 439 GRVEMEMEHMAERIAVLAQDLVR 465
DB 643 GRKXVEMTKQMYKRFVDVLYKCMR 669

```

RESULT 6  
ABB91331  
ID ABB91331 standard; Protein, 670 AA.

```

XX AC ABB91331;
XX DT 31-MAY-2002 (first entry)
XX DE Herbicidally active polypeptide SEQ ID NO 542.
XX KM Herbicidal; plant; agriculture; herbicide.
XX OS Arabidopsis thaliana.
XX

```

```

XX PN WO00210210-A2.
XX PD 07-FEB-2002.
XX PF 28-AUG-2001; 2001WO-EP09892.
XX PR 28-AUG-2001; 2001WO-EP09892.
XX PA (FARB ) BAYER AG.
XX PI Tjerjen K, Weidler M;
XX DR WPI; 2002-269010/31.
XX PT Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms -
XX
XX PS Claim 5; SEQ ID NO 542; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
XX Sequence 670 AA;
XX
Query Match 20.1%; Score 482.5; DB 23; Length 670;
Best Local Similarity 29.2%; Pred. No. 2.1e-40;
Matches 148; Conservative 77; Mismatches 173; Indels 109; Gaps 13;
QY 38 STAGFLIR--GALRD-----PDGRGDPALMTAVAGSPUGFMR--SKLYV 80
DB 193 ATYGKLLGPFGLSEKYLEWSPHRRSGTCDKRS-----FKRLVMSRRV 237
QY 81 LVSHSLSSGPELLMELAFLLRHVSQVWITNQRQETNDVTYSLERHMLNHYOVLP 140
DB 238 LRLHELSMTGARISMELSELSCGATSAVNLSTRG-----GLMGELSRRIWV 290
QY 141 ARQGEAVDIALKADVLILNTAVAGKMLDVKDHVPKYLKILWIMHMKGHYK----- 195
DB 291 DKGLSFTKAMKADLILAGSAVCTSWIDQYNNHH-PAGGSQIAWIMENRRREYFDRAPV 349
QY 196 VEYVKHLPPVAGAMIDSHHTAEYWNH-RTSRLKIQMPQTYVVLGNSKEIMEVA----- 249
DB 350 LDRVKKMLPL-----SESQSRQWLWCEEBHILKLR-SQVILVPLSVDELAFAGIPSS 402
QY 250 -----EDNVARVLREHRESLGVRSSELLFALINVSAGKQDLFLQATYQALQ- 299
DB 403 LMTPLTSPBKRVKVKQILRESVTELGITDSMLWMSLSINFQGLLLESTALALSB 462
QY 300 -----LIQHEKLPRIHAH-----VYGS 318
DB 463 RQGESQRNHKGIIRKEKYSLSKGRLRGSSQMKSVSLTLNGLRREKQELKVLGVS 522
QY 319 DVNAQCKFETQLRDFVVKNTIHRVYFVKVTLAAVFLAIDLVLQNSQSGEGEFGRTI 378
DB 523 KSNKGVYKEMLSPLSNSGNLSKSVMTPTATFRVAAVSLYSAADVYVNSQGVGTFEGVYI 582
QY 379 EAAAFKLPVLTAGGTTEVLIDGSTGLHPAGKEGVAFLAKNIVRLASHAEQFVSNGEK 438
DB 583 EAAAYGLAVGTGAGTKEWQHNMTGLHSGSGSKKELAHNLVYLRLNDELRLRGSE 642
QY 439 GYGRVYKEMPEHMAERLAAVLKDVLK 465
DB 643 GRQGVKQMKYKQHKYKRFVDVLVKCR 669

```

```

RESULT 7
ABB57933
ID ABB57933 standard; Protein; 424 AA.
XX AC ABB57933;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 591.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO000171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers BW;
XX DR WPI; 2001-656860/75.
XX DR N-PSDB; ABL02036.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Disclosure; SEQ ID NO 591; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC sequences (AB101840-AB116175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 424 AA;
XX
Query Match 7.5%; Score 179.5; DB 22; Length 424;
Best Local Similarity 22.2%; Pred. No. 2.1e-09;
Matches 101; Conservative 73; Mismatches 181; Indels 99; Gaps 15;
QY 79 VLVSHSLSSGPELLMELAFLLRHVSQVWITNQRS-----QETNDVTYSLERHMLN 133
DB 4 VLFALPDLGIGGAERLVVDALALKEKGHVSLTNHDSHCKEFLADGTQPV----- 57
QY 134 HGVOVLPRQGEAVDIALKADVLILNTAVAGK-----LDPLKDHVPKYLKILWMI 186
DB 58 HVGAGWLPRGLGFRFYALICAYLRMLYAATYASFMPQREQVVDVVCGLSVCP-----V 112
QY 187 HEMRGHYEYKYLPPVAGAMIDSHHTAEYWNHRTSDRLKIQMPQTYVVLHG-----N 241
DB 113 LRFAPHRKVLFCFCHP-----DQLSSRSGLLKRLYRLPIWLEBHITGLADKVLVN 165
QY 242 SKELMEVAEDNVAR-----RVLR-----EHRESLGVRSSELL-----LF 275
DB 166 SKFTLRVQDTEFRRLSTVPDVLVPSLHTQYFDQMKQKLEQSSALLDPEVPRVPLNAFTY 225

```



QY 276 AIINVSRRGKQDLFLQAFYQALQIQHEKLVPRIHAVVVG-----SDVNA 322  
 DB 226 LDIRRYRRKKNAHALHLSRLKGLMDLPATDFK--RCRLIINGGVDTRCMENVEHAELEH 283  
 QY 323 QTKETQLRDFV-----KNTIHDRVHFVNKTLAVPYLAIDVLVONSQSGECFGRIT 377  
 DB 284 LIE-ELKIQDHVVLLRSPTDEKCRLLFAHCLLYTP-----ENERHGIIVP 328  
 QY 378 IEAFAFLPYLGTAAAGTTEIVLDGSTGLHPAGKEGVAFLAKNIIVLASHAEORVSMGE 437  
 DB 329 LEGVYCKSPVVALNSGPTETVSTSGPLCEKTEKS---FGAMQLPFDQGLVKRGD 385  
 QY 438 KGYGRVKEMEMEHMAERIALAVLDVLRKSQEH 471  
 DB 386 QGHKRVQOKFSPQAFADRLNGIIRDVLVPIRESS 419

RESULT 8  
 AAY68959  
 ID AAY68959 standard; Protein; 385 AA.  
 AC AAY68959;  
 DT 30-MAY-2000 (first entry)  
 XX Cps2G protein which has glycosyltransferase activity.  
 DE Cps2G  
 XX Capsular gene cluster; serotype 2; polysaccharide biosynthesis;  
 KM capsular component; antigen regulation; chain length determination;  
 KM complement-mediated opsonophagocytosis; serotype-specific detection;  
 KM antigen; vaccine; Streptococcal disease; ORF 22; ORF 21; ORF 22;  
 KM Cps2A; Cps2B; Cps2C; Cps2D; Cps2E; Cps2F; Cps2G; Cps2H; Cps2I; Cps2J;  
 KM Cps2K; Cps2O; Cps2P; Cps2Q; Cps2R; Cps2S; Cps2T.  
 XX Streptococcus suis.  
 OS  
 XX  
 FH Key  
 FT Misc-difference 266 /note="unspecified amino acid encoded by AGA"  
 FT  
 XX W0200005378-A2.  
 PN 03-FEB-2000.  
 PD 19-JUL-1999; 99WO-NL00460.  
 PP 22-JUL-1998; 98EP-0202465.  
 PR 22-JUL-1998; 98EP-0202467.  
 XX  
 PA (DIEN-) STICHTING DIENST LANDBOUWKUNDIG ONDERZOE.  
 XX  
 PI Smith HE;  
 XX  
 XX MPI: 2000-195104/17.  
 DR N-PSDB; AA260929.  
 XX  
 PT New nucleic acid containing the capsular gene cluster of Streptococcus  
 PT suis, used for serotype-specific detection and to generate antigens or  
 PT mutants for vaccination  
 XX  
 PS Disclosure; Fig 3; 144pp; English.

CC humans or animals, e.g. against S. suis in pigs.  
 XX  
 SQ Sequence 385 AA;  
 Query Match 6.7%; Score 161.5; DB 21; Length 385;  
 Best Local Similarity 24.4%; Pred. No. 1.3e-07;  
 Matches 90; Conservative 61; Mismatches 145; Indels 73; Gaps 20;

QY 102 LRVGSQVVMITNO--RSQETN-----DVTYSLEHRLMHGVQLPARGQEAVDIALKAD 154  
 DB 48 LREVGQVVEIVINYPILRRKYFNPKGIFDYFISYHH---YSKOI---AQYALIE--NKVD 97  
 QY 155 LVLIINT--AVAGKMLDVLKDNHVKTLPKILMWIHEMRGHYFVEYVK--HPFVAGAMID 211  
 DB 98 ITHNNTAVLEGIVYKRLK-----LP-LIMHVEH---IVPKFISDSINFLMGFFAD 147  
 QY 212 SHTTAEVMSRTSDLK-----IQMPQTVVHLGSKMEVAEDVARRVLRHIRESLG 267  
 DB 148 KIVTV--SQAVANHIKOSPHIKDDISVIYGVNKFVYQSD---ARSV-----RERFD 196  
 QY 268 VRSEDLIPALINVSRRGKQDLFLQAFYQALQIQHEKLVPRIHAVVVGSDVNAQ----- 323  
 DB 197 IDEBALVIGMGRVANMKGQDFLEAVAPILE-----QNPKALIFLAGAFEGEEMRV 249  
 QY 324 TKFETQLRDFVVKNTIHDRVHFVNKTLAVPYLAIDVLVONSQSGECFGRITTEMAF 383  
 DB 250 VELEKKSQKISQVSKMDYVANTT---ELVNMFDIPVLPST--NPDLPTVYLKAMAC 304  
 QY 384 KLPLYGTAAAGTTEIVLDGSTGLHPAGKEGVA-----LAKNIIVLASHAEORVSMGEK 438  
 DB 305 GKPVVGIRGIGCEVKEGVNGFL-----VTPSPPLNLSKVIQLQSENINLKRKIGNN 357  
 QY 439 GYGRVKEMF 447  
 DB 358 SIEROKEHF 366

RESULT 9  
 AAG43119  
 ID AAG43119 standard; Protein; 405 AA.  
 AC AAG43119;  
 DT 18-OCT-2000 (first entry)  
 XX Arabidopsis thaliana protein fragment SEQ ID NO: 53657.  
 DE Arabidopsis thaliana  
 XX Protein identification; signal transduction pathway; metabolic pathway;  
 KM hybridisation assay; genetic mapping; gene expression control; promoter;  
 KM termination sequence.  
 XX Arabidopsis thaliana.  
 OS  
 XX EP1033405-A2.  
 PN 06-SEP-2000.  
 PD 25-FEB-2000; 2000EP-0301439.  
 PP 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 25-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0133048.  
PR 04-MAY-1999; 99US-0133407.  
PR 05-MAY-1999; 99US-0133484.  
PR 06-MAY-1999; 99US-0133485.  
PR 06-MAY-1999; 99US-0133486.  
PR 07-MAY-1999; 99US-0133487.  
PR 11-MAY-1999; 99US-0133486.  
PR 14-MAY-1999; 99US-0134215.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 18-MAY-1999; 99US-0133370.  
PR 18-MAY-1999; 99US-0133768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138847.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139452.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140822.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142330.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145129.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 05-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151530.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155133.  
PR 23-SEP-1999; 99US-0155386.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.

PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

Query Match 6.7%; Score 161; DB 21; Length 405;

Best Local Similarity 22.3%; Pred. No. 1.6e-07;

Matches 97; Conservative 74; Mismatches 186; Indels 78; Gaps 17;

QY 76 SKLVLVSH-ELISGGPILMELAFILRHVSGOV-VWITNQRSGQETNDVYSLERMLN 133  
DB 7 SKMIAIITHPDIGAEERLIVDAAVELASHGKHHTSHHDSRCFEETISGIFQVTV 66  
QY 134 HGVOVLPR---GQEAVDIALKADLVILNTAVAGKMDPVLKDHVFKVLP-----KI 182  
DB 67 YG-SFLPRHIFYRLHVAVCAYLCLFVALCVLGMSSFDVLADVSVPVELKLRSSKV 125  
QY 183 LWMIH-----EMRGHYFK-VEYVGHLPFVAGAMI--DSHTTAEYMSRTSDRLK 228  
DB 126 VPCYHPDILLAKHTTLRLRMVRRKPIDFIEQGTGMAIMLVNPNFASGF-ANTFKRLN 184  
QY 229 IQMPTVYVHLSKSKLEVAEDNVARVLRHIRESLGVRSEDLALINVSRSKQOD 288  
DB 185 AGRSPAVLYPA-----VNIQDFLEPHYKLANLSINRERKNIID 225  
QY 289 LFLQFYQALQIHEKLVKVPRIHAVVVG-----SPVNAQTKRFETQLRDPVVKTIHDRV 343  
DB 226 LAVSAF---AIIICKKQKLSVTLTVAAGCGYDEKLENVEYLELRSLAEKEGVSDRV 281  
QY 344 HFV-----NKTILAVPYLAIDVIVONSQRCGECRRTITAMAFKLPVLTAGGT 395  
DB 282 NFITSCSTARNELLS-----SCLCVLYTPD---EHFELVLEMAAYEPVIAACNSGAP 333  
QY 396 TEIVLDGSGTG-LHPAGKEGVAPLAKNIVLASHAQRVSGEKGVRVEMEHMAE 454  
DB 334 VETVANGVTGLCEPPEDEFSAMV---RFIENPELANMGAEARRHNVVESFSVYTGQ 389  
QY 455 RIAMVTKDVKRSQE 469  
DB 390 KUNQYLVVVVSSPE 404

RESULT 10  
AAG43120  
ID AAG43120 standard; Protein; 397 AA.  
XX  
AC AAG43120;  
DE 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 53858.  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
OS Arabidopsis thaliana.

XX  
PN EP1033405-AZ.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127452.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0134256.  
PR 11-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0138119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.



OY 459 VLKDLKKSQE 469  
 Db 386 YLVNVSSPE 396

RESULT 11  
 AAG24139  
 ID AAG24139 standard; Protein; 403 AA.

AC AAG24139;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 27699.

KM Protein identification; signal transduction pathway; metabolic pathway;  
 KM hybridisation assay; genetic mapping; gene expression control; promoter;  
 KM termination sequence.

OS Arabidopsis thaliana.

PN EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 28-APR-1999; 99US-0130891.  
 PR 30-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 04-MAY-1999; 99US-0132407.  
 PR 05-MAY-1999; 99US-0132484.  
 PR 06-MAY-1999; 99US-0132485.  
 PR 07-MAY-1999; 99US-0132486.  
 PR 07-MAY-1999; 99US-0132487.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134319.  
 PR 14-MAY-1999; 99US-0134321.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139750.  
 PR 18-JUN-1999; 99US-0139763.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139899.  
 PR 23-JUN-1999; 99US-0140353.  
 PR 24-JUN-1999; 99US-0140354.  
 PR 28-JUN-1999; 99US-0140695.  
 PR 29-JUN-1999; 99US-0140823.  
 PR 30-JUN-1999; 99US-0140991.  
 PR 01-JUL-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142058.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142803.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144006.  
 PR 16-JUL-1999; 99US-0144088.  
 PR 16-JUL-1999; 99US-0144088.  
 PR 19-JUL-1999; 99US-0144326.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 20-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0144814.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 21-JUL-1999; 99US-0145088.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 27-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145951.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.

PR	13-AUG-1999;	99US-014684.
PR	16-AUG-1999;	99US-0149366.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149472.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153756.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155446.
PR	24-SEP-1999;	99US-0155658.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158363.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159359.
PR	14-OCT-1999;	99US-0159350.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	6.5%;	Score 157;	DB 21;	Length 403;
Best Local Similarity	22.2%;	Pred. No. 4.2e-07;		
Matches 96;	Conservative 74;	Mismatches 187;	Indels 76;	Gaps 17

[illegible]

```

Db      67 YG-SLEPRIRIPIRLAAVAVIATLCFLVATCIVLGMSSFDVLADQVSVPEBLKLRSSKV 125
Qy      183 LAMWIT-----EMRGHYFK-VEYVKHLRPAVAGMT--DSHTAEYWNRSQSDPLK 228
      ::::
Db      126 VFCHPREDLLAKHTTTLRRMYRKPIDIEEOTTGNADMTLVNNSFNPASTF-ANFKFLYN 184
Qy      229 IOWPOTVVVHLGNSKEIMEVAEDNVARVRLEHRIRESLQVRSBELLPAIINSVSRGKOD 268
      ::::
Db      185 AQSRAVLVYRA-----VNIDGFIRPHYKLVNPLNSINPFERKKIID 225
Qy      289 LFLQAPYQALDIOHEKTKYRRIHAAVWVG--SDVAQCKREKQEDLPYVVKVTTIDRYNF 345
      ::::
Db      226 LANSARF---AIIKCKONISDVTLTVAAGYDRLKENVETIEELRSLEAKESVDSDRNF 261
Qy      346 V-----NKTLLAVRYALDVLVWNSGQGECECFGRITTEAMAFRLPYLGTAAGSTTE 397
      ::::
Db      282 ITSCTAERNELTA-----SCLCVLYRPTD--EHFGIYVPELMAAAKRYIACNSGGGEVE 333
Qy      388 IYVDSGSTG-LTHPRGKEGVAPLAKNIYALASHAEQVSVSGKEGVRKUEFMENHMAEI 456
      ::::
Db      334 TVKNGVTVGLDEPRPEPSSAMA-----RLENELLANRGAELANNHVESFVCKTFQGL 369
Qy      457 AAIVLCKVLRKSOE 469
      ::::
Db      390 NQYLVVWVSSPE 402
      ::::

```

RESULT 12	ID	SEQ	ID NO:
AAAG24140	AAAG24140	standard; Protein; 395 AA.	27700
XX	XX		
AC	AAAG24140;		
XX	XX		
DT	17-OCT-2000	(first entry)	
XX	XX		
KE	Arabidopsis thaliana protein fragment		

OS Arabidopsis thaliana.  
KM Protein identification; signal transduction pathway; metabolic pathway;  
KM Proben identification assay; genetic mapping; gene expression control; promoter  
XX termination sequence.

PR	25-FEB-1999;	99US-0121825
PR	05-MAR-1999;	99US-0123180
PR	09-MAR-1999;	99US-0123548
PR	23-MAR-1999;	99US-0125788
PR	25-MAR-1999;	99US-0126284
PR	29-MAR-1999;	99US-0127685
PR	1-APR-1999;	99US-0127462
PR	06-APR-1999;	99US-0128824
PR	08-APR-1999;	99US-0128774
PR	16-APR-1999;	99US-0129845
PR	19-APR-1999;	99US-0130077
PR	21-APR-1999;	99US-0130510
PR	23-APR-1999;	99US-0130981
PR	28-APR-1999;	99US-0131449
PR	30-APR-1999;	99US-0132048
PR	3-APR-1999;	99US-0132407
PR	04-MAY-1999;	99US-0132484
PR	05-MAY-1999;	99US-0132485
PR	06-MAY-1999;	99US-0132486
PR	06-MAY-1999;	99US-0132487
PR	07-MAY-1999;	99US-0132863
PR	11-MAY-1999;	99US-0134256
PR	14-MAY-1999;	99US-0134218
XX	25-FEB-2000;	2000EP-0301493

PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136382.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139891.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145115.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 03-AUG-1999; 99US-0146389.  
PR 04-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151067.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152365.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155488.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0158293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161362.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161922.  
PR 28-OCT-1999; 99US-0161932.  
PR 28-OCT-1999; 99US-0161933.  
PR 29-OCT-1999; 99US-0162142.

Query Match 6.4%; Score 153.5; DB 21; Length 395;  
Best Local Similarity 21.7%; Pred. No. 9.3e-07;

Matches 93; Conservative 74; Mismatches 187; Indels 75; Gaps 16;

QY 79 VLVSHELSSGGLPMLMELAFLLRHVGSQV-WITNORSQETNDVTVSLERHMLNMQV 137  
DB 3 IAIHFDLIGCAERLIYDAVAELASHGHKVIIFSHHDSRCFETLSGIFQTVYG-S 61  
QY 138 VLPR---GQAVDIALKADVLITNVAAGKMLDPVKDHPVLP-----KILMT 186  
DB 62 FLRHHFYRLHAXVAYLRCLFVALCVLGMSSFDVLADQSVVFLKLRSSKVVFC 121  
QY 187 H-----EMRGHYFK-VBYVNHLPVAGAMT--DSHTTAEYMSRTSDCLKIQMP 232  
DB 122 HFPDLLAKHTTTLRRMYRKRPIDEEQTGMADMIIVNSNFTASTP-ANTFKELNAGS 180  
QY 233 QTVVHLGNSKELMEVAEDNVARVLRREHRESIGVSEDLPAIINSVSRGKQDLFIQ 292  
DB 181 RPAVLPA-----VNIQFIERTTYLNTLSINFEKKKIDILANS 221  
QY 293 AFYQALQIQHEKLVPRHIAVVG---SDVNAQTEKTEQLRDFVKNQTHDRVHEV--- 346  
DB 222 AF---AIIQCKKONLSDVTLTVAGGYDERLKENVEYELBELRLAEKGVSDRNFITSC 277  
QY 347 -----NKTIAVAPYLAIDVLVONSQGECEGFCRITTEAFAFLPYLGAAGTTEIVLD 401  
DB 278 STABNELLS-----SCICVLYTPTD---EHGIVPLEMAAKVPIACNSGGEVETVKN 329  
QY 402 GSGG-LIHPAGKEGVAFLAKNIYRLASHAQRVSNGEKGIVKEMEFEMHMERIANVYL 460  
DB 330 GVTGYICEPTPEDSSMAA---RLENEELANRNGAEARNHVVSFSVTKFGOKNQYL 385  
QY 461 KDVLRSQDE 469  
DB 386 VDVVSSPKE 394

RESULT 13

AA93231  
ID AA93231 standard; Protein; 368 AA.

AA93231;

04-SEP-2000 (first entry)

An Escherichia coli virulence protein.

Virulence protein; catA, catB, catC, catE, mdgG, creC, recG, y99N;

Gram negative bacterium.

Escherichia coli.

MO200028038-A2.

18-MAY-2000.

09-NOV-1999; 99WO-GB03721.

09-NOV-1998; 98GB-0024569.

09-NOV-1998; 98GB-0024570.

17-DEC-1998; 98GB-0027814.

PR 17-DEC-1998; 98GB-0027815.  
PR 17-DEC-1998; 98GB-0027816.  
PR 17-DEC-1998; 98GB-0027818.  
PR 13-JAN-1999; 99GB-0000708.  
PR 13-JAN-1999; 99GB-0000710.  
PR 13-JAN-1999; 99GB-0000711.  
PR 28-JAN-1999; 99GB-0001915.

(MICR-) MICROSCIENCE LTD.

Crooke HR, Clarke EE, Everest PH, Dougan G, Holden DW, Shea JE;

Fieldman RG;

WPI; 2000-376550/32.

N-PSDB; AAA15182.

Peptide encoded by an operon including genes from Escherichia coli for

screening potential drugs, detecting virulence and treating conditions

associated with infection by a Gram negative bacterium -

Claim 2; Page 66-67; 122pp; English.

The present sequence represents an Escherichia coli virulence protein.

The specification describes virulence proteins which are encoded

by an operon including catA, catB, catC, catE, mdgG, creC, recG, y99N,

catK, iroD, iroC, iroE, mtd2 or msl-16 genes obtained from Escherichia

coli K1. The virulence proteins and polynucleotides, and their vaccines

are useful for screening potential drugs, for the detection of virulence,

and for treating or preventing conditions associated with infection by

a Gram negative bacterium particularly Escherichia coli.

Sequence 368 AA;

Query Match 6.4%; Score 152.5; DB 21; Length 368;

Best Local Similarity 21.7%; Pred. No. 1.1e-06;

Matches 90; Conservative 68; Mismatches 150; Indels 107; Gaps 19;

QY 65 VASGSPFGFMRKSLVLLVSHLSLGGPMLMELAFLLRHVGSQVWITNORSQETNDVT 124  
DB 6 VITGIGLG-GAERQVOLLADKLSLGHVVKIISLGM-----SNKRV 46  
QY 125 YSLERHMLNHGVOLPARQGEAVDIALKADVLITNVAAGKMLDPVKDHPVLPKILM 184  
DB 47 FPSE-----NNVNVIVVNSKNISGVKGCVR-----RDVIANFPDIL-- 85  
QY 185 WHEMRGHYKREYVNHLPFA-----GAMDSHTTAE--YMSRT-----SD----- 225  
DB 86 ---VSHMFPNANITITRLSVIGINRPGIISTAHNNKNGYFRMLTYRTDCLSDCTNV 141  
QY 226 -----RLKIOMPOTVVHLGNSKELMEVAEDNVARVLRREHRESIGVSEDLFLFA 276  
DB 142 SKEAVDEFIRKAFNP-AKAITMYNGIDTNKPKFLLARR-----EINDGINKNDIILL 196  
QY 277. IINSVRGKQODFLQAFYQALQIQHEKLVPRHIAVVGSDVNAQTEKTEQLRDPD--- 332  
DB 197 AAGRLTIAKQYENLHAM---TLRPHFKL-----ITIG-----DGLBDEIIM 237  
QY 333 FVVKNTIHDRVHFVNKTLAVALPYLAIDVLVONSQGECEGFCRITTEAFAFLPYLGA 392  
DB 238 LTKIQLSNVSVLGVKNIAPFSACDIPVLS--RWEGGLVVAEAMSCRRIVGIDS 295  
QY 393 GGTTEIVLDSSTGLIHPAGKEGVAFLAKNIYRLASHAQRVSNGEKGIVKEMF 447  
DB 296 GGVRREVI--GDDDFLVFISDS--TQLASKIEKL-SLSQIRDHIFRNRERILKNF 345

RESULT 14

ABB52953  
ID ABB52953 standard; Protein; 368 AA.

ABB52953;

11-FEB-2002 (first entry)



XX Escherichia coli polypeptide SEQ ID NO 1322.

XX Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;  
KW immunosuppressive; extra-intestinal infection; phylogeneity; meningitis;  
KW systemic infection; non-diarrhoeal infection; septicemia;  
KW pyelonephritis; antibiotic resistance.

XX Escherichia coli.

XX W0200166572-A2.

XX 13-SEP-2001.

XX 12-MAR-2001; 2001WO-EP03445.

XX 10-MAR-2000; 2000FR-0003145.  
PR 02-FEB-2001; 2001FR-0001449.

XX (INRM) INSEEM INST NAT SANTE & RECH MEDICALE.

PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;

DR WPI; 2001-550253/61.

PT A library of DNA fragments of Escherichia coli strains for the  
PT phylogenetic determination of a given strain comprises polynucleotides of  
PT nature B2/D+ A- -

XX Example 6; Fig 6; 646pp; English.

XX The invention relates to a library of DNA fragments of Escherichia coli  
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)  
CC and encoded proteins (ABB52459-ABB52919) of nature B2/D+A-. The  
CC polynucleotides have potential antiinflammatory, antibacterial and  
CC immunosuppressive activity as part of pharmaceutical compositions used to  
CC treat, palliate or prevent extra-intestinal E. coli infections. The  
CC polypeptides are useful for determining the phylogenetic group of a given  
CC E. coli strain. These polypeptides can detect and treat an undesired  
CC development of E. coli, particularly an extra-intestinal infection that  
CC include systemic and non-diarrhoeal infections such as septicemia,  
CC pyelonephritis and meningitis this is particularly advantageous as  
CC bacterial resistance is increasing with the more frequent use of broad  
CC spectrum antibiotics.

XX Sequence 368 AA;

XX Query Match 6.4%; Score 152.5; DB 22; Length 368;

XX Best Local Similarity 21.7%; Pred. No. 1.1e-06;

XX Matches 90; Conservative 68; Mismatches 150; Indels 107; Gaps 19;

XX 65 VASGSPGLGFMRSKLVLYVSHLSLSGGPILLMELAFILRHVGSQVWITNQSQTNDVT 124

XX 6 VITGIGLG-GAEKQVCLADKLKSLSGHHVKTLSLGHM-----SNNKV 46

XX 125 YSLERHMLNMGVQVLPARGOAVDIALKADLVILNTAVAGKMLDPVLCBHVPCVLPKILM 184

XX 47 FFSSE-----NNVVNIIVNMNSKNISGVIKGCVR-----RDVIANFKPDI-- 85

XX 185 WIEHRGHYFVKEYVKLPLFVA-----GAMIDSHTTAA--YVNSRT-----SD----- 225

XX 86 -----VHSHMFANITITLSVIGIKORPGIISTANNKGGYRMLTYITDCLSDCCNTV 141

XX 226 -----RLKIOMPTVYVHLGSKELMVAEDNVARVRAHRIIRSLGVSSEDLTA 276

XX 142 SKEAVDEPLRTKAFNP-AKAITMYGIDTNKFKFDLARR-----EIRGINIKNDITLL 196

XX 277 IINSVSRGQDPLFAFYQALQTLQHEKLVKPRIHAVVGSVDAVNAQKTFETQLRD----- 332

XX 197 AAGRLTLAKDYPNLLNM--TLPEHFFL-----IIIG-----DGEIRDEINM 237

XX 333 FVVKQTHIDRHVFNPKTLAAVRYTLAIDVLVNSQSGSECFRITTEAMAKLPVGTGA 392

Db 238 LIKKQLSNRVSLLGVKKNIAFPYSACDIFVLS--RMEGGLVVAEMSCERTIVGTS 295

Qy 393 GGTTEIVLDGTLGHPAGKEGVAPLAKNIVLASHAQRVSGKGYGVKEMF 447

Db 296 GGRFVFI--GDDDFIVRIDS--TQASKIEKL-SLQQLNDHIGFRNNRILTKNF 345

RESULT 15

ABB52987

ABB52987 standard; Protein; 368 AA.

XX ABB52987;

XX 11-FEB-2002 (first entry)

XX Escherichia coli polypeptide SEQ ID NO 1322.

XX Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;  
KW immunosuppressive; extra-intestinal infection; phylogeneity; meningitis;  
KW systemic infection; non-diarrhoeal infection; septicemia;  
KW pyelonephritis; antibiotic resistance.

XX Escherichia coli.

XX W0200166572-A2.

XX 13-SEP-2001.

XX 12-MAR-2001; 2001WO-EP03445.

XX 10-MAR-2000; 2000FR-0003145.  
PR 02-FEB-2001; 2001FR-0001449.

XX (INRM) INSEEM INST NAT SANTE & RECH MEDICALE.

PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;

DR WPI; 2001-550253/61.

PT A library of DNA fragments of Escherichia coli strains for the  
PT phylogenetic determination of a given strain comprises polynucleotides of  
PT nature B2/D+ A- -

XX Example 6; Fig 6; 646pp; English.

XX The invention relates to a library of DNA fragments of Escherichia coli  
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)  
CC and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature  
CC B2/D+A-. The polynucleotides have potential antiinflammatory,  
CC antibacterial and immunosuppressive activity as part of pharmaceutical  
CC compositions used to treat, palliate or prevent extra-intestinal E. coli  
CC infections. The polypeptides are useful for determining the phylogenetic  
CC group of a given E. coli strain. These polypeptides can detect and treat  
CC an undesired development of E. coli, particularly an extra-intestinal  
CC infection that include systemic and non-diarrhoeal infections such as  
CC septicemia, pyelonephritis and meningitis this is particularly  
CC advantageous as bacterial resistance is increasing with the more  
CC frequent use of broad spectrum antibiotics.

XX Sequence 368 AA;

XX Query Match 6.4%; Score 152.5; DB 22; Length 368;

XX Best Local Similarity 21.7%; Pred. No. 1.1e-06;

XX Matches 90; Conservative 68; Mismatches 150; Indels 107; Gaps 19;

XX 65 VASGSPGLGFMRSKLVLYVSHLSLSGGPILLMELAFILRHVGSQVWITNQSQTNDVT 124

XX 6 VITGIGLG-GAEKQVCLADKLKSLSGHHVKTLSLGHM-----SNNKV 46

XX 125 YSLERHMLNMGVQVLPARGOAVDIALKADLVILNTAVAGKMLDPVLCBHVPCVLPKILM 184

XX 47 FFSSE-----NNVVNIIVNMNSKNISGVIKGCVR-----RDVIANFKPDI-- 85

Qy 185 WIHEMRGYFKVEYVGHLPVA-----GAMIDSHTTAE--YVNSRT-----SD----- 225  
Db 86 ----VHSHMFHANIITRLSVIGIKNRPGLISTAHKXKMGYFRLITYRITDCLSDCCTNV 141  
Qy 226 -----RLKIOMPQTYVHLGNSKELMEVAEDNVARVLRHHIRESLGVRESEDLFA 276  
Db 142 SKEAVDEFLRIKAFNP-AKAITWYNGIDTNKFKFDLLAR---EIRDGINIKNDILL 196  
Qy 277 IINSVSRKGODLFLQAFYQALQIOHEKLVPRIHAVVQSDVNAQTKEETOLRD----- 332  
Db 197 AAGRLTLAKDYPNLNNM---TLPEHFKL-----IITG-----DGLRDEINM 237  
Qy 333 FVVYNTIHDRVHFVNKTIHAVPYLALDVLVONSQGECEGRITTEAMAFKLPVLTAA 392  
Db 238 LIRKLQSNRVSLLGVKKNIAPYFSACDIFVLSS--RWEGFGLVVAEAMSCERIVGTDS 295  
Qy 393 GGTTETIVLDGSTGLHPAGKEGVAPLAKNIVRLASHAEORVSMGEGYGVKEMF 447  
Db 296 GGVREVI--GDDDFLVPISDS--TQLASKIERL-SLQIRDHIGFPRRERILKNF 345

Search completed: June 30, 2003, 16:26:51  
Job time : 74 secs

20/ GAMUDSHI IAE MNSKRSIDUKENLQVFI VV - NUSNOXDLDFEVRNUNVWVAV  
| : | : : | : : | : : | : : |

Db 132 -----KLIBFSLFLDKFILLISEXYLIANHISFNKSFSLINNGV--EVITGDSRN 182  
Qy 261 HIRESLGVSEDLFAIINSVSRGKQDPLQAFYQALQLOHEKLUKPRHNAVYGSV 320  
Db 183 EIEEIEF--PNEDFTIOMVGRUSPPPEFFFDPAKKLQIQRNDTNF-----ITVGGG- 232  
Qy 321 NAQTKETQLRDPVKNVTIHDRVHFVNKTLAVAPYLAIDVLVONSQGRGECFGRIT-E 379  
Db 233 ---ELRSEIERMILNDGLGDKIYITGWDNPNRYIEKFDQALIFSREGL---SLTIAE 285  
Qy 380 ANAFKLPVLTAGTTEIYLDGSTGLHPAGKEGVAFLAKNIYRLASHAEQVSMGEKG 439  
Db 286 YMSQKKTILATINIGINDLITDGEFTGLIEVGDLSA-VSKSP-ELRNNKVSNOQLANNA 343  
Qy 440 YGRVKEMF-MEHMAE 454  
Db 344 YNKVVEQFSIEKQMAE 359

## RESULT 2

US-08-746-682A-8  
Sequence 8, Application US/08746682A  
Patent No. 5786184  
GENERAL INFORMATION:  
APPLICANT: STINGELE, Francesca  
APPLICANT: MOULET, Beat  
TITLE OF INVENTION: LACTIC BACTERIA PRODUCING  
TITLE OF INVENTION: EXOPOLYSACCHARIDES  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americans  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/746, 682A  
FILING DATE: 14-NOV-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/597, 236  
FILING DATE: 20-JUN-1995  
APPLICATION NUMBER: EP 95201669.9  
FILING DATE: 20-JUN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Rannucci A., Allen  
REGISTRATION NUMBER: 30256  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 372 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-746-682A-8

Query Match 5.2%; Score 125.5; DB 1; Length 372;  
Best Local Similarity 21.8%; Pred. No. 0.00013;  
Matches 69; Conservative 62; Mismatches 134; Indels 51; Gaps 17;

Qy 151 LKADVLVINTAVAGKLDVPLKDHVPKV-LP-KILMWTHERGHEVFE--YVKLPLFVA 206  
Db 83 INPIVVLHSTFAG-----VGRIASIGLPTKVYTNPH---GSPFMDYSTKTLIF-- 131  
Qy 207 GAMIDSTTAEYWNSTSDRLKTIOMPTQYV--HLGNSKEIMEVAEDNVARREV----RE 260  
Db 132 -----KLIBFSLFLDKFILLISEXYLIANHISFNKSFSLINNGV--EVITGDSRN 182

Qy 261 HIRESLGVSEDLFAIINSVSRGKQDPLQAFYQALQLOHEKLUKPRHNAVYGSV 320  
Db 183 EIEEIEF--PNEDFTIOMVGRUSPPPEFFFDPAKKLQIQRNDTNF-----ITVGGG- 232  
Qy 321 NAQTKETQLRDPVKNVTIHDRVHFVNKTLAVAPYLAIDVLVONSQGRGECFGRIT-E 379  
Db 233 ---ELRSEIERMILNDGLGDKIYITGWDNPNRYIEKFDQALIFSREGL---SLTIAE 285  
Qy 380 ANAFKLPVLTAGTTEIYLDGSTGLHPAGKEGVAFLAKNIYRLASHAEQVSMGEKG 439  
Db 286 YMSQKKTILATINIGINDLITDGEFTGLIEVGDLSA-VSKSP-ELRNNKVSNOQLANNA 343  
Qy 440 YGRVKEMF-MEHMAE 454  
Db 344 YNKVVEQFSIEKQMAE 359

## RESULT 3

US-09-134-001C-3206  
Sequence 3206, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134, 001C  
PRIOR FILING DATE: 1998-08-13  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/064, 964  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3206  
LENGTH: 387  
TYPE: PRN  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3206

Query Match 5.0%; Score 119; DB 4; Length 387;  
Best Local Similarity 18.7%; Pred. No. 0.00069;  
Matches 80; Conservative 77; Mismatches 173; Indels 98; Gaps 15;

Qy 87 SLSSGFLIMELAFILRHVSQVWVITNQRSGETNDVYSLHRMNL---HGOYVLARG 143  
Db 17 SMGSSGIATELGIMARHEVHFI-----TSNIPRIKRLPNMTFHOVEV----N 65  
Qy 144 QEAV-----DIALKADVLINTAVAGKLDVPLKDHVPKV-LKILMWTHER 188  
Db 66 QYAVFOYPPYDITLSTKISDVIOEYDLILHMYA-----VPIAVCGIL--AKQ 112  
Qy 189 MRGHVEVEYVGLPVPVAGAMID-----SHTTAEVWNSRTSDRLK 228  
Db 113 MSGGNVKIMTTLHGHDITVIGYDHTLOANIKFGIEQSDIYTSVSHLQ----- 161  
Qy 229 IQMPQTYVHLGNSKEIMEV---AEDNVARRYLREHIRESLGVSEDLFAIINSVSRGK 285  
Db 162 ---QYVEI-INTKKEIIPYVNFRENEPPTRHNEELKDCYISPEKVLHVSFNPKYK 216  
Qy 286 GODLFLQAFYQALQLOHEKLUKPRHNAVYGSVDNAQTGFTQLRDPVKNVTIHDRVHF 345  
Db 217 RIDPVIETFAKV-----HESLP-----SKLILGDPPELLIMRKARELDVET---HVL 263  
Qy 346 VNKTLAVAPYLAIDVLVONSQGRGECFGRITIAVAFAKLPVLTAGTTEIYLDGSTG 405  
Db 264 LGRQNDVSAFYQLSDVLVLLSE--KESFGLTLLPAMKTGVLPISGHAIGKEVIRHEETG 321  
Qy 406 LHPAGKEGVAFLAKNIYRLASHAEQVSMGEKGYGRVKEMFMEHMAERIAAVLKQVLR 465  
Db 322 FIVYDIDSTQA--AKVAIKLSNPPELYQKMSQMLKQIARRSLSLIDQYENYRKMLE 379  
Qy 466 KSGEHSRS 473

Db 380 OGENNES 387

RESULT 4  
US-09-403-768-8  
Sequence 8, Application US/09403768

Patent No. 644804  
GENERAL INFORMATION:  
APPLICANT: Iam Joseph S.  
APPLICANT: De Kievit, Teresa R.  
APPLICANT: Burrows, Lori L.  
APPLICANT: Walsh, Andrew  
APPLICANT: Matewish, Mauricia  
TITLE OF INVENTION: No. 644804e1 Proteins Involved in the Synthesis and Assembly  
FILE OF INVENTION: of Core Lipopolysaccharide of Pseudomonas Aëringinosa  
FILE REFERENCE: 6580-177  
CURRENT APPLICATION NUMBER: US/09/403,768  
PRIOR FILING DATE: 1997-05-02  
PRIOR APPLICATION NUMBER: U.S. 60/046,149  
PRIOR FILING DATE: 1997-05-09  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 8  
LENGTH: 373  
TYPE: PR1  
ORGANISM: Pseudomonas aeruginosa Rfag (Maag)  
FEATURE:  
OTHER INFORMATION: Description of Organism: Pseudomonas aeruginosa  
US-09-403-768-8

Query Match 4.5%; Score 108.5; DB 4; Length 373;  
Best Local Similarity 20.9%; Pred. No. 0.0083;  
Matches 48; Conservative 39; Mismatches 98; Indels 45; Gaps 8;

QY 228 KIQMPQYVYHLGNSKEIMEVADNVA--RRV-----LREHRESLGVSDDLFAIN 279  
DB 145 EVQOP-LLVKHVGTQAEHFLLPFGISQDRAPANAADVAFRRRFGLEEDLLVQIG 203  
QY 280 SVSNGKODLFLQAFYQALQIHEKLVPRIHAVVGSVDNAQTFFETQLDEPVNQT- 338  
DB 204 SGRTKGDRLSKALSLPKALRR-----FRLIAIGD-----DPKFLIQIAA 248  
QY 339 --IHDRVHVNTKTLAVAPYLAIDVYVNSQSGCFRITTEMAFKLPVGTGAGTT 396  
DB 249 LGINDQVQILKGRSDIPFLIGADLIHPAY--NENTGTVLLELVSGLPVLTVDGGA 306  
QY 397 EIVLDGSTGLHPAGKE-----GVAFLAKNIVRLASHAE 430  
DB 307 HYIADAGKVLPSFFEDSLNRLLEMLEDAPARAASRGGLAVDHAD 356

RESULT 5  
US-08-941-445A-5  
Sequence 5, Application US/08941445A

Patent No. 6107060  
GENERAL INFORMATION:  
APPLICANT: Keeling, Peter  
APPLICANT: Guan, Hanning  
TITLE OF INVENTION: Starch Encapsulation  
NUMBER OF SEQUENCES: 37  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Greenlee, Winner and Sullivan, P.C.  
STREET: 5370 Manhattan Circle  
CITY: Boulder  
STATE: CO  
COUNTRY: US  
ZIP: 80303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/941,445A  
FILING DATE: 30-SEP-1997  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/026,855  
FILING DATE: 30-SEP-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Winner, Ellen P  
REGISTRATION NUMBER: 28,547  
REFERENCE/DOCKET NUMBER: 89-97  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (303) 499-8080  
TELEFAX: (303) 499-8089  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 533 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-941-445A-5

Query Match 4.5%; Score 108.5; DB 3; Length 533;  
Best Local Similarity 22.8%; Pred. No. 0.015;  
Matches 44; Conservative 35; Mismatches 73; Indels 41; Gaps 7;

QY 254 ARVRREHIRESLG--VRESDLLFAINSVSRGKODLFLQAFYQALQIHEKLVPR1 311  
DB 302 AKALNKEKQAEVGLPYDNIPLVAFICRLEBQKGPVWAAAIPLMEWED-----V 354  
QY 312 HAVVGSVDNAQIKFETQLRDFVKNTHDRVHFVNKTLAVAPYLA1-----DULV 363  
DB 355 QIVLLGT--GKKKFERMLMSAEK-----FPKQVAVVFNALAHHINAGADVIA 403  
QY 364 QNSQSGCFRITTEMAFKLPVGTGAGTTEIVLDSTG-----LHPACK 412  
DB 404 VHS--RFPQCLIQDQMRGTGPCACASTGSLVDITIEGTFHNGRLSDVCNVEPADV 461  
QY 413 EGVAPLAKNIVRL 425  
DB 462 KKVATTLQRAIKV 474

RESULT 6

US-08-861-464-10  
Sequence 10, Application US/08861464  
Patent No. 5874210

GENERAL INFORMATION:  
APPLICANT: Guarente, Leonard P.  
APPLICANT: Austriaco Jr., Nicanor  
TITLE OF INVENTION: Genes Determining Cellular Senescence  
NUMBER OF SEQUENCES: 16  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive  
CITY: Lexington  
STATE: MA  
COUNTRY: USA  
ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/861,464  
FILING DATE: 22-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

```

/ APPLICATION NUMBER: US 08/396,001
/ FILING DATE: 28-FEB-1995
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US94/09351
/ FILING DATE: 15-AUG-1994
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/107,408
/ FILING DATE: 16-AUG-1993
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Granahan, Patricia
/ REGISTRATION NUMBER: 32,227
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 781-861-6240
/ TELEFAX: 781-861-9540
/ INFORMATION FOR SEQ ID NO: 10:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 508 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-861-464-10

```

```

Query Match          4.4%; Score 105.5; DB 2; Length 508;
Best Local Similarity 21.4%; Pred. No. 0.029;
Matches 97; Conservative 75; Mismatches 140; Indels 141; Gaps 26;

```

```

QY 101 LIRHVGSOVV-----WITNRSQETNDV---TYSL-----EHRMLNHGVVLPARGOE 145
DB 105 MLRHAESAIVEAYVNDKAILLEORMLTEELYGNTFOLYSADHRTLDKYLEVQPEKTEL 164
QY 146 AVD-----IALKADLVILNTAVAGKMLD-----PVLKDHVPKVLPRILMIHEMRGH 192
DB 165 IMDEMKQILTPMAOK-EAVIKSHLVKVFDFFTYAP-----PKLRSEMTBAIRE----- 213
QY 193 YFKVEYVHLRPFVAGAMIDSHTTAAYWNSRTSDR-LKIQMPQTYVHVGNSK--ELMEVA 249
DB 214 --AVVYLATH--DGRVAMHC---LHGTPEKDKRVIVKTMKTYEKAANGQYSHLVILA 266
QY 250 -----EDNVARVULREHIRESLGVRSBDLFAIINSVSRGQDLFLQAFYQALQLOH 303
DB 267 AFDCIDDTLVKQIILISEISLP-----SIVND-KYGRKVLILSPDPRAHTVR- 316
QY 304 EKLKVRIRHVVGVSDVNAQTKEFTQLRDVFVKNTHDRVHFVNKTLAVAPYLA--AIDV 361
DB 317 -----ELIEVLQKDGNAHSHKQTEVRRELESI-----SPALLSYLOEHAQEV 361
QY 362 LVNQSOGREBCGFRITTEMAFKLPVLGTAGGT---TEIYLDGSTGLHPACKRGVAP 417
DB 362 VLDSK-----ACVLVSDILGSATGVQPTMNAISLAATG-LHFGKDGELH 407
QY 418 LAKNIVRLASH-----AEQVSMGEGK-YGRVKEMFMEH----- 450
DB 408 IAEH--PAGHLVLKWLIEQDKKMKENGREGCFAKTLVEHVGKMKLKSMAVNRGAILLS 464
QY 451 -----HMAERIAAVLKDV---LRKQSHSR 472
DB 465 SLIQSCDLEVANKVAKLSLIPLETKTSTSK 497

```

```

RESULT 7
US-08-396-001-10
/ Sequence 10, Application US/08396001
/ Patent No. 5919618
/ GENERAL INFORMATION:
/ APPLICANT: Guarente, Leonard P.
/ APPLICANT: Austriaco Jr., Nicanor
/ APPLICANT: Claus, James
/ APPLICANT: Cole, Francesca
/ APPLICANT: Kennedy, Brian
/ TITLE OF INVENTION: Genes Determining Cellular Senescence in
/ YEAST
/ NUMBER OF SEQUENCES: 16

```

```

/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
/ STREET: Two Wiltia Drive
/ CITY: Lexington
/ STATE: MA
/ COUNTRY: USA
/ ZIP: 02173
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/396,001
/ FILING DATE: 28-FEB-1995
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Granahan, Patricia
/ REGISTRATION NUMBER: 32,227
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 617-861-6240
/ TELEFAX: 617-861-9540
/ INFORMATION FOR SEQ ID NO: 10:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 508 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-396-001-10

```

```

Query Match          4.4%; Score 105.5; DB 2; Length 508;
Best Local Similarity 21.4%; Pred. No. 0.029;
Matches 97; Conservative 75; Mismatches 140; Indels 141; Gaps 26;

```

```

QY 101 LIRHVGSOVV-----WITNRSQETNDV---TYSL-----EHRMLNHGVVLPARGOE 145
DB 105 MLRHAESAIVEAYVNDKAILLEORMLTEELYGNTFOLYSADHRTLDKYLEVQPEKTEL 164
QY 146 AVD-----IALKADLVILNTAVAGKMLD-----PVLKDHVPKVLPRILMIHEMRGH 192
DB 165 IMDEMKQILTPMAOK-EAVIKSHLVKVFDFFTYAP-----PKLRSEMTBAIRE----- 213
QY 193 YFKVEYVHLRPFVAGAMIDSHTTAAYWNSRTSDR-LKIQMPQTYVHVGNSK--ELMEVA 249
DB 214 --AVVYLATH--DGRVAMHC---LHGTPEKDKRVIVKTMKTYEKAANGQYSHLVILA 266
QY 250 -----EDNVARVULREHIRESLGVRSBDLFAIINSVSRGQDLFLQAFYQALQLOH 303
DB 267 AFDCIDDTLVKQIILISEISLP-----SIVND-KYGRKVLILSPDPRAHTVR- 316
QY 304 EKLKVRIRHVVGVSDVNAQTKEFTQLRDVFVKNTHDRVHFVNKTLAVAPYLA--AIDV 361
DB 317 -----ELIEVLQKDGNAHSHKQTEVRRELESI-----SPALLSYLOEHAQEV 361
QY 362 LVNQSOGREBCGFRITTEMAFKLPVLGTAGGT---TEIYLDGSTGLHPACKRGVAP 417
DB 362 VLDSK-----ACVLVSDILGSATGVQPTMNAISLAATG-LHFGKDGELH 407
QY 418 LAKNIVRLASH-----AEQVSMGEGK-YGRVKEMFMEH----- 450
DB 408 IAEH--PAGHLVLKWLIEQDKKMKENGREGCFAKTLVEHVGKMKLKSMAVNRGAILLS 464
QY 451 -----HMAERIAAVLKDV---LRKQSHSR 472
DB 465 SLIQSCDLEVANKVAKLSLIPLETKTSTSK 497

```

```

RESULT 8
US-09-323-433A-10
/ Sequence 10, Application US/09323433A
/ Patent No. 6218512
/ GENERAL INFORMATION:

```

```

/ APPLICANT: Guarente, Leonard P.
/ APPLICANT: Austriaco Jr., Nicnor
/ APPLICANT: Claus, James J.
/ APPLICANT: Cole, Francesca
/ APPLICANT: Kennedy, Brian
/ TITLE OF INVENTION: GENES DETERMINING CELLULAR SENESENCE IN
/ TITLE OF INVENTION: YEAST
/ FILE REFERENCE: 0050.1491-003
/ CURRENT APPLICATION NUMBER: US/09/323,433A
/ CURRENT FILING DATE: 1999-06-01
/ PRIOR APPLICATION NUMBER: US 08/396,001
/ PRIOR FILING DATE: 1995-02-28
/ PRIOR APPLICATION NUMBER: PCT/US94/09351
/ PRIOR FILING DATE: 1994-08-15
/ PRIOR APPLICATION NUMBER: US 08/107,408
/ PRIOR FILING DATE: 1993-08-16
/ NUMBER OF SEQ ID NOS: 48
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 10
/ LENGTH: 508
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-09-323-433A-10

```

```

Query Match      4.4%; Score 105.5; DB 4; Length 508;
Best Local Similarity 21.4%; Pred. No. 0.029;
Matches 97; Conservative 75; Mismatches 140; Indels 141; Gaps 26;

```

```

QY 101 LHHVSGQV-----WITNRSQETNDV---TYSL-----EHRMLNHGVLPARGOE 145
DB 105 MLHRAASAIIEYAYNDKALIEQRNMLTELYGTFOLYVSADHRLTKVLEVOEDEL 164
QY 146 AVD-----IALKADLVILNTAVAGKWD-----PYLKHVPKVLPRILMWIHEMRGH 192
DB 165 IMEMKQIILTPAQK-EAVIKSHLVKVFLEDFEYAP-----PKLRSEMIKIRE----- 213
QY 193 YFKYEVYKHLPFVAAAMIDSHTTAAYMNSRTSDR-LKIQMPQYVYVHLGNSK--ELMEVA 249
DB 214 --AVVYLAHTH--DARVAMHC--LMHGTPDKRVKVKTKYVERKANNGYSHLVLA 266
QY 250 -----EDNVARRVLRHIRESLGVSEDLPAIINSVSRGKODLFLQAFYALQIQH 303
DB 267 AFDCIDDTKLVKQIILSIISLP-----SIYND-KYGRKVLILYLSPPDPANHTVR- 316
QY 304 EKLKVRPIHAAVVGSDVNAQKTEPTOLRDPVVKNTIHRRVAFNNKTLAAVPLA--ALDV 361
DB 317 -----ELIEVLQKGDGNASHKKOTEYRRREILSEI-----SPALLSTLQEHQEV 361
QY 362 LVONSQGRGECFGRITIEBMAFKLPVLGTAAGT-----TEIVLDGSTGLHPAGKEGVAP 417
DB 362 VLDKS-----ACVIVSDILGSATGDVQPTNMAIASIAATG-LHPGCKDGLH 407
QY 418 LAKNIVTLASH-----AEGRVSMGEKG-YGRVKEMPEH----- 450
DB 408 IAEH--PAGHLVTLKWLIEQKKMKENGECFPAKTIVHEVGNLKSMAVNRGAIIIS 464
QY 451 -----HMAERIAVAVLKDV--LRKSOEHR 472
DB 465 SLQSCDLVAVANKVKAALKSLPLTEKTKSTK 497

```

```

RESULT 9
US-09-335-409-5
/ Sequence 5, Application US/09335409
/ Patent No. 6121029
/ GENERAL INFORMATION:
/ APPLICANT: Schupp, Thomas
/ APPLICANT: Ligon, James
/ APPLICANT: Molnar, Istvan
/ APPLICANT: Zirkle, Ross
/ APPLICANT: Cyr, Devon
/ APPLICANT: Goerlach, Joern
/ TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIOLONES

```

```

/ FILE REFERENCE: 4-30582A
/ CURRENT APPLICATION NUMBER: US/09/335,409
/ CURRENT FILING DATE: 1999-06-17
/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 5
/ LENGTH: 7257
/ TYPE: PRT
/ ORGANISM: Sorangium cellulosum
/ US-09-335-409-5

```

```

Query Match      4.4%; Score 104.5; DB 3; Length 7257;
Best Local Similarity 19.7%; Pred. No. 3.2;
Matches 100; Conservative 61; Mismatches 146; Indels 201; Gaps 21;

```

```

QY 58 PALNTAVASGSPGPMRSKVLVLSHELISGPLLMEIAF-----LHRVSGQ 108
DB 6623 PALNTAAQAAALPVAEMTWGLV--HIGRLRAGEVLIHSYTGTTGLAAVQIARHLGAE 6680
QY 109 V-----WITNRSQETNDVYSLSEHRMLNHGVLPARGOEAVDIALKADLVIL 158
DB 6681 IFRTAGTEPERKAMLRD-----GLAHVMSRSLDFAEQVLAAATGEGVD-----VVL 6727
QY 159 NTAVAGKWLDPVLDKQHPKVLKILMWIHEMRGHYFRVEYVKNLPFVAGAMDSHTAEY 218
DB 6728 NS-----LSGALIDA----- 6737
QY 219 VNSRTSDRLKIQMPQYVYVHLGNSKELMEVAEDNVARRVL-REHIRESLGVSESD----- 272
DB 6738 -----SLSTLVDPDGRFIEKLT-----DIYDRSLGLAHFRKSLSYSAVDLAGLA 6782
QY 273 -----LPAIINSVSRGKODLFLQAF-----YQALQIQH----- 303
DB 6783 VRPERVALLAEVVDLARGALQPLPELIFLPSRAADAFRMAQOGLKVLALDEDD 6842
QY 304 EKLKVRPIHAAVVGSDVNAQKTEPTOLRDPVVKNTIHRRVAFNNKTLAAVPLA----- 357
DB 6843 VRIKVPDSSGVALRAD-----GALVYTGIGL-----GLISVAGMLAEGGAGH 6885
QY 358 -----ALDVLYONSQGRGECFGRITIE-----EMAKFLPVYGT-- 390
DB 6886 LVIVGRSGAVSAEQOTVAALLENHGRVTVARADVADPRQMERIIRREYTAGNGLKRVVH 6945
QY 391 AAGGTTEIVLDGST-----GLHPAGKEGVAPLAKNIVTLASHAEGRVSMGE 437
DB 6946 AAGIIDDGLMQVPPARFRAVVA PKVRGALHLHLTREAPLS-FVLVYASGAGILSPQ 7004
QY 438 KGYGRVK--EMPEHMAERIAVAVLKD 462
DB 7005 GNYAANNTFLDALHRRRAOGGLPALSID 7032

```

```

RESULT 10
US-09-568-102-5
/ Sequence 5, Application US/09568102
/ Patent No. 6346404
/ GENERAL INFORMATION:
/ APPLICANT: Schupp, Thomas
/ APPLICANT: Ligon, James
/ APPLICANT: Molnar, Istvan
/ APPLICANT: Zirkle, Ross
/ APPLICANT: Cyr, Devon
/ APPLICANT: Goerlach, Joern
/ TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHIOLONES
/ FILE REFERENCE: 4-30582A
/ CURRENT APPLICATION NUMBER: US/09/568,102
/ CURRENT FILING DATE: 2000-05-10
/ PRIOR APPLICATION NUMBER: 09/335,409
/ PRIOR FILING DATE: 1999-06-17
/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 5
/ LENGTH: 7257

```





```
Db 6681 IFATAGTPEKRAMLRQ-----GIAHWDSRSLDFAEVLATKGEVD-----VVL 6727
Qy 159 NTAVAGKMLDVLKDNHVKPLPKILMWHMERGHYFVVEVGHLPFVAGMIDSHHTAY 218
Db 6728 NS-----LSGAALIDA----- 6737
Qy 219 WNSRTSDRLKIOMQOTVYVHLGNSKELMEVADNVARVL-REHIRESLGVRSED----- 272
Db 6738 -----SLSTLVPDGFIELGKT-----DIYADRSIGLAHFRKSLYSYSAVDLAGLA 6782
Qy 273 -----LLFAINSVRKGODLFLQAF-----YQALQLOIOLH----- 303
Db 6783 VRPRERVAALLAEVVDLLARGALQPLVEIFPLSRADAFRKMAOHLGKLVLALEDDP 6842
Qy 304 EKLKVRPIHAAVVGSDVNAQTKFETQLRDFVKNKTIHDRVHFVNKTLAAVPIA----- 357
Db 6843 VIRIVGESGVALRAD-----GAYLVGTGLG-----GLGLSVAGWLAEQAGH 6885
Qy 358 -----AIDVLVONSQGRGECFG-RITI-----EAMAFKLPVLGT-- 390
Db 6886 LVLVGSAGVSAEQOTVAALLAEHAGARTVVARADVADRAQMERILREVTASGMPLRGVH 6945
Qy 391 AAGCTTEIVLDGST-----GLHPAGEGVAPLAKNIVPLASHAEQVSMGE 437
Db 6946 AAGIIDDGLMOQTPARFRAVMAPKVKGALHHLALTREAPLS-FFVLVYASGAGLLSPQ 7004
Qy 438 KGYGRVK---EMPEHMAERIAVLKD 462
Db 7005 GNYAANTFIDLALHRRAGLPLALSID 7032

RESULT 13
US-09-568-486-5
; Sequence 5, Application US/09568486
; Patent No. 6355459
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligot, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,486
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-486-5

Query Match 4.4%; Score 104.5; DB 4; Length 7257;
Best Local Similarity 19.7%; Pred. No. 3.2;
Matches 100; Conservative 61; Mismatches 146; Indels 201; Gaps 21;

Qy 58 PALNTAVASGSPGLGFMRSKVLVLVSHLSLGGPLIMELAF-----LIRHVSGQ 108
Db 6623 PALTLTAQAALPFAFTMYGLV--HLGRLRAGERVLISHATGCTGLAAVQARHLGAE 6680
Qy 109 V-----WITNORSQETNDVYVLSLHRMLAHGVQVLPARGQENVDAIKADLVIL 158
Db 6681 IFATAGTPEKRAMLRQ-----GIAHWDSRSLDFAEVLATKGEVD-----VVL 6727
Qy 159 NTAVAGKMLDVLKDNHVKPLPKILMWHMERGHYFVVEVGHLPFVAGMIDSHHTAY 218
Db 6728 NS-----LSGAALIDA----- 6737
Qy 219 WNSRTSDRLKIOMQOTVYVHLGNSKELMEVADNVARVL-REHIRESLGVRSED----- 272
Db 6738 -----SLSTLVPDGFIELGKT-----DIYADRSIGLAHFRKSLYSYSAVDLAGLA 6782
Qy 273 -----LLFAINSVRKGODLFLQAF-----YQALQLOIOLH----- 303
Db 6783 VRPRERVAALLAEVVDLLARGALQPLVEIFPLSRADAFRKMAOHLGKLVLALEDDP 6842
Qy 304 EKLKVRPIHAAVVGSDVNAQTKFETQLRDFVKNKTIHDRVHFVNKTLAAVPIA----- 357
Db 6843 VIRIVGESGVALRAD-----GAYLVGTGLG-----GLGLSVAGWLAEQAGH 6885
Qy 358 -----AIDVLVONSQGRGECFG-RITI-----EAMAFKLPVLGT-- 390
Db 6886 LVLVGSAGVSAEQOTVAALLAEHAGARTVVARADVADRAQMERILREVTASGMPLRGVH 6945
Qy 391 AAGCTTEIVLDGST-----GLHPAGEGVAPLAKNIVPLASHAEQVSMGE 437
Db 6946 AAGIIDDGLMOQTPARFRAVMAPKVKGALHHLALTREAPLS-FFVLVYASGAGLLSPQ 7004
Qy 438 KGYGRVK---EMPEHMAERIAVLKD 462
Db 7005 GNYAANTFIDLALHRRAGLPLALSID 7032
```

```
Db 6738 -----SLSTLVPDGFIELGKT-----DIYADRSIGLAHFRKSLYSYSAVDLAGLA 6782
Qy 273 -----LLFAINSVRKGODLFLQAF-----YQALQLOIOLH----- 303
Db 6783 VRPRERVAALLAEVVDLLARGALQPLVEIFPLSRADAFRKMAOHLGKLVLALEDDP 6842
Qy 304 EKLKVRPIHAAVVGSDVNAQTKFETQLRDFVKNKTIHDRVHFVNKTLAAVPIA----- 357
Db 6843 VIRIVGESGVALRAD-----GAYLVGTGLG-----GLGLSVAGWLAEQAGH 6885
Qy 358 -----AIDVLVONSQGRGECFG-RITI-----EAMAFKLPVLGT-- 390
Db 6886 LVLVGSAGVSAEQOTVAALLAEHAGARTVVARADVADRAQMERILREVTASGMPLRGVH 6945
Qy 391 AAGCTTEIVLDGST-----GLHPAGEGVAPLAKNIVPLASHAEQVSMGE 437
Db 6946 AAGIIDDGLMOQTPARFRAVMAPKVKGALHHLALTREAPLS-FFVLVYASGAGLLSPQ 7004
Qy 438 KGYGRVK---EMPEHMAERIAVLKD 462
Db 7005 GNYAANTFIDLALHRRAGLPLALSID 7032

RESULT 14
US-09-568-472-5
; Sequence 5, Application US/09568472
; Patent No. 6358719
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligot, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/568,472
; PRIOR FILING DATE: 2000-05-10
; PRIOR APPLICATION NUMBER: 09/335,409
; PRIOR FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 7257
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-568-472-5

Query Match 4.4%; Score 104.5; DB 4; Length 7257;
Best Local Similarity 19.7%; Pred. No. 3.2;
Matches 100; Conservative 61; Mismatches 146; Indels 201; Gaps 21;

Qy 58 PALNTAVASGSPGLGFMRSKVLVLVSHLSLGGPLIMELAF-----LIRHVSGQ 108
Db 6623 PALTLTAQAALPFAFTMYGLV--HLGRLRAGERVLISHATGCTGLAAVQARHLGAE 6680
Qy 109 V-----WITNORSQETNDVYVLSLHRMLAHGVQVLPARGQENVDAIKADLVIL 158
Db 6681 IFATAGTPEKRAMLRQ-----GIAHWDSRSLDFAEVLATKGEVD-----VVL 6727
Qy 159 NTAVAGKMLDVLKDNHVKPLPKILMWHMERGHYFVVEVGHLPFVAGMIDSHHTAY 218
Db 6728 NS-----LSGAALIDA----- 6737
Qy 219 WNSRTSDRLKIOMQOTVYVHLGNSKELMEVADNVARVL-REHIRESLGVRSED----- 272
Db 6738 -----SLSTLVPDGFIELGKT-----DIYADRSIGLAHFRKSLYSYSAVDLAGLA 6782
Qy 273 -----LLFAINSVRKGODLFLQAF-----YQALQLOIOLH----- 303
Db 6783 VRPRERVAALLAEVVDLLARGALQPLVEIFPLSRADAFRKMAOHLGKLVLALEDDP 6842
```

**UB-09-938-294-45.rai**

```

QY      304  EKATPRHIAVWVGSDNQTKTEFQJDRDVVKNIIHNRHNFVNKTLAVAPYLA----- 357
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      6643  VRIRPBGSSGALTAAD-----GAYLVTTGGLG-----GLGTSVAGWLLEQGAGH 6885
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      358  -----AIDVLVONSOGKGECEFG-RITL-----EMAFKLPVLGT-- 390
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      6886  LVLVSGSAGVSAEQRTAAVLALEAHGAVVTYARADVADRAQMERILREVTYASGMPLGUVH 6945
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      391  AAGGTTEVLVDGSI-----GLIHPRAGEVAPLAPLAKNIYRLASHAEQVWSMGE 437
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      6946  AAGTIDDELMOQTPARFPAVAPKVGALHILHALTREAPLS-FFVLYASAGALGPSQ 7004
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      438  KGYGRVK---EMFENHMAERIALVLD 462
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      7005  GATYAAMTFEDALAHHRRAQGLPDLSTID 7032

```

```

QY      391  AAGGTTETVLDSGR-----GLTHPAGKEGVAPLAKNIYTRLSAHNEQRYVME 437
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      6946  AAGILDDGLMQOTPARPRVWAPKVGALHHLHATLRAPLS-FFLVYASGAGLLSPQ 7004
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY      438  KKGGRVK---EMEMEHMKERLAALVKD 462
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db      7005  GNYAAMNTFDLALHHRRKQGGLPALSTD 7032
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

```

RESULT 15
US-09-567-899-5
: Sequence 5, Application US/09567899
: Patent No. 6383787
: GENERAL INFORMATION:
: APPLICANT: Schnupp, Thomas
: APPLICANT: Ligon, James
: APPLICANT: Molnar, Istvan
: APPLICANT: Zitzke, Rose
: APPLICANT: Cyr, Devon
: APPLICANT: Goerlach, Joern
: TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF BOTRIOLONES
: FILE REFERENCE: 4-30582A
: CURRENT APPLICATION NUMBER: US/09/567,899
: CURRENT FILING DATE: 2000-05-10
: PRIOR APPLICATION NUMBER: 09/335,409
: PRIOR FILING DATE: 1999-06-17
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 5
: LENGTH: 7257
: TYPE: PRT
: ORGANISM: Sorangium cellulosum
: US-09-567-899-5

```

Query Match	4.4%;	Score 104.5;	DB 4;	Length 7257;
Best Local Similarity	19.7%;	Pred. No. 3.2;	Mismatches 146;	Indels 201;
Matches	100;	Conservative	61;	Gaps 21.

QY	58	PAALNTAVAGSGPFGMRSKIVLLVSHLSISGCEPLLMLBELA----	-----LIRHVSGQ 108
DB	6623	PAALTPAOMALPVAEMTWAGLV- <u>HLGRL</u> AGERVILSHSAGTGTGLAAVQIARHLGA	6680
QY	109	V-----VWITNRSQETNDVYLSLEHMLHNGVQLPARGEANDTALAADIYVL	158
DB	6681	IPATGTPERKAMIRGQ-----GIAIWDNSLDPFAOYLAATKKEGD-----VVL	6727
QY	159	NTAVAGKWLDPVLKDHPVLPKILMWIHMRGHYFKVEYVKKLPVGAAMIDSHTTAEY	218
DB	6728	NS-----LSGAIDA-----	6737
QY	219	WMSRTSDRLKIQMPQTYVYVHLGNSKEIMEVAEDNARVL-RRHITRESIGVRS	272
DB	6738	-----SLSTVLPDRFLEIGKT-----DIYARSLGIAHFRKSIYSAYVDLAGHA	6782
QY	273	-----LLFAIINSVSRGKGODLFLQAF-----YQALQILQH-----	303
DB	6783	VRRPERVAALLAEVVDLARGLDPLPVELFELSPAADAFRRKAQAOHLGKLYALIEDPD	6842
QY	304	EKLAKPIRIHAVVVGSDVNAQTKEFTQLRDFVVKNTIHDRVHFNKTKLVAEYLA-----	357
DB	6843	VARIRVPSGVAIRAD-----GAVLVTGILG-----GIGLSVAGMALDEGAGH	6885
QY	358	-----AIDVLVNSGGRGECFG-RITL-----EAMAKPLVTCG-----	390
DB	6886	LVLVGRSGAVSAEQCTAAVALLAEHGAIRVAVADADVADPAOMERILREVTASGMPFLGVGH	6945

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 30, 2003, 16:29:51 ; Search time 51 Seconds

(without alignments)  
1016.921 Million cell updates/sec

Title: US-09-938-294-45

Perfect score: 2397

Sequence: 1 MAKTPSFVAVAAGSGRPVH.....ERIAVVKDVLKSGOHSRS 473

## Scoring table:

BLOSUM62

Searched: 424699 seqs, 109646833 residues

Total number of hits satisfying chosen parameters: 424699

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/2/pubppaa/PTCT\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB pep.\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB pep.\*  
7: /cgn2\_6/ptodata/2/pubppaa/CTTUS\_PUBCOMB pep.\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUB pep.\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB pep.\*  
10: /cgn2\_6/ptodata/2/pubppaa/US10\_NEW\_PUB pep.\*  
11: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB pep.\*  
12: /cgn2\_6/ptodata/2/pubppaa/US10\_PUBCOMB pep.\*  
13: /cgn2\_6/ptodata/2/pubppaa/US60\_NEW\_PUB pep.\*  
14: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	181	7.6	191	10	US-09-924-358-22
2	179	7.5	144	10	US-09-924-358-25
3	162.5	6.8	385	10	US-09-767-041-18
4	150	6.3	387	9	US-09-738-626-3890
5	134.5	5.6	416	10	US-09-924-358-5
6	128.5	5.4	409	9	US-09-738-626-4738
7	128.5	5.4	418	9	US-09-738-626-3951
8	127	5.3	381	9	US-09-738-626-5896
9	126.5	5.3	416	10	US-09-815-028-5
10	118	4.3	371	9	US-09-870-759-107
11	113	4.7	720	9	US-10-217-700-14
12	111	4.6	323	9	US-10-174-590-296
13	111	4.6	323	9	US-10-176-758-296
14	111	4.6	323	9	US-10-175-737-296
15	111	4.6	323	9	US-10-173-706-296
16	111	4.6	323	9	US-10-175-738-296
17	111	4.6	323	9	US-10-175-752-296
18	111	4.6	323	9	US-10-176-482-296
19	111	4.6	323	9	US-10-176-757-296

20	111	4.6	323	9	US-10-176-913-296	Sequence 296, App
21	111	4.6	323	9	US-10-180-552-296	Sequence 296, App
22	111	4.6	323	9	US-10-180-557-296	Sequence 296, App
23	111	4.6	323	9	US-10-173-700-296	Sequence 296, App
24	111	4.6	323	9	US-10-174-572-296	Sequence 296, App
25	111	4.6	323	9	US-10-174-579-296	Sequence 296, App
26	111	4.6	323	9	US-10-174-588-296	Sequence 296, App
27	111	4.6	323	9	US-10-174-588-296	Sequence 296, App
28	111	4.6	323	9	US-10-175-739-296	Sequence 296, App
29	111	4.6	323	9	US-10-175-740-296	Sequence 296, App
30	111	4.6	323	9	US-10-175-743-296	Sequence 296, App
31	111	4.6	323	9	US-10-176-488-296	Sequence 296, App
32	111	4.6	323	9	US-10-176-492-296	Sequence 296, App
33	111	4.6	323	9	US-10-176-747-296	Sequence 296, App
34	111	4.6	323	9	US-10-176-750-296	Sequence 296, App
35	111	4.6	323	9	US-10-176-985-296	Sequence 296, App
36	111	4.6	323	9	US-10-176-987-296	Sequence 296, App
37	111	4.6	323	9	US-10-176-991-296	Sequence 296, App
38	111	4.6	323	9	US-10-176-992-296	Sequence 296, App
39	111	4.6	323	9	US-10-176-993-296	Sequence 296, App
40	111	4.6	323	9	US-10-184-658-296	Sequence 296, App
41	111	4.6	323	9	US-10-173-695-296	Sequence 296, App
42	111	4.6	323	9	US-10-173-697-296	Sequence 296, App
43	111	4.6	323	9	US-10-173-705-296	Sequence 296, App
44	111	4.6	323	9	US-10-174-576-296	Sequence 296, App
45	111	4.6	323	9	US-10-174-585-296	Sequence 296, App

## ALIGNMENTS

RESULT 1  
US-09-924-358-22  
; Sequence 22, Application US/09924358  
; Patent No. US20020107376A1  
; GENERAL INFORMATION:  
; APPLICANT: Millennium Pharmaceuticals, Inc.  
; APPLICANT: Meyers, Rachel  
; TITLE OF INVENTION: 26199, 33530, 33949, 47148, 50226, AND  
; TITLE OF INVENTION: 58764  
; FILE REFERENCE: 38155-20034.00  
; CURRENT FILING DATE: 2001-08-06  
; PRIOR APPLICATION NUMBER: US/09/924,358  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 22  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Consensus amino acid  
US-09-924-358-22

Query Match 7.6% Score 181; DB 10; Length 191;  
Best Local Similarity 32.7% Pred. No. 1e-08; Indels 24; Gaps 9;  
Matches 65; Conservative 32; Mismatches 78;  
259 REHRESIGVRESDLLFAIINSVRKGGDPLQAFYQALQIQHEKLVPIIHAVVGS 318  
2 REEIRKKLGIGKDKKIIIFVGRILVPEKIGDILLIEAFKKKKPKLRLKNL-PVILKLVYGG 60  
QY DVNAQT-KPEITLQRFVVKNTIHDRVAFNKTILAVP-----YLAATDVVQNSQGRG 370  
Db PYDESDDEEDDEKTLAKLGLLENVIF----LGFVPEDDLPELYKSA-DVPLVS--RY 113  
QY ECFGRITTEAMAFKPLVIGT-AMCGTTEIVLDGSTGLLHPAGKGVAPLAKIYVLTASHA 429  
Db EGFGIVLEMMACGPIVATNCVGIPEVVKDGETGLLVERGDPEA-LAEAIETKLKDE 172



```
Db 115 VRIIVYVHAGRSTETVLGKKKIVLEVLVQL-FANQTLAVSHLKKALIEDHAPFKR 173
Qy 227 LKIQMPQYVVHLC---NSKELMEVADNARAVLEHEI--RESIGRSDDLFAIINS 280
Db 174 VQV-----LGYGSMNSVEL-----DRPRVSLBEKLSANMLNIPSSKSYGVGR 219
Qy 281 VSRGKGODFLQAFYQALQLOIHEKLVRIHAAVVGSDVNAQTKEFTOLRDPVVK--- 336
Db 220 INKDKGDDL-----AALTKEAFTRLRHLHLIIE-----LEDDLEAFIKLVNE 266
Qy 337 --NTIHDVHFVNKTLAVAPYLAIDVLYONSQGRGCFRTITIEAMFKLPVIGTAGG 394
Db 267 GQVTLTGWIDPPEEP-----LAAVDVLIHPLOREG--LQMSLLEQMGVPLTNATVG 318
Qy 395 TTEIVLDGSTGLHPAGKEGAPLANIVRLASHAEORVSGEKGYGKEMFEMHHMAE 454
Db 319 TVDAVYSGEGGF--ADDDSVESWVSKIDLVSPKLRDMGRAGROFVSARFVRDVA 376
Qy 455 RIA 457
Db 377 RFS 379

RESULT 5
US-09-924-358-5
; Sequence 5, Application US/09924358
; Patent No. US20020107376A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; APPLICANT: Macbeth, Kyle
; TITLE OF INVENTION: 26199, 33530, 33949, 47148, 50226, AND
; TITLE OF INVENTION: 58764,
; TITLE OF INVENTION: NOVEL HUMAN TRANSFRASE FAMILY MEMBERS AND USES THEREFOR
; FILE REFERENCE: 38155-2003.4.00
; CURRENT APPLICATION NUMBER: US/09/924,358
; PRIOR FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: US 60/229,300
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 416
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-924-358-5

Query Match 5.6%; Score 134.5; DB 10; Length 416;
Best Local Similarity 21.7%; Pred. No. 0.00062;
Matches 92; Conservative 61; Mismatches 193; Indels 77; Gaps 17;

Qy 79 VILVSHELISGSPILLMEALFLLRHVGSQV-VWITNQRSGETNDVTYSLERMLNHGVQ 137
Db 17 VLFHLPDLGVGHERLVLDLALALQARGCSVKIWTAR-----YRPGCFRPS--R 64
Qy 138 VLPR-----GQEARVDIALKADIVILNTAV--AGKMLDVLKDHVPKVL- 180
Db 65 ELPYRCAGDWLPRGLMGWGRGAACAYRVAFALYVFLADEFEFDVWDDOVSACIPVF 124
Qy 181 -----KILMWI-----EMRGHYFKEYVYVGHLPFVAGAMIDSHTTAEVWNSRSTDR 227
Db 125 RLARRRKKILFYCHPFDLLTKRDSPLKRIYRPIDW-----IEVYTG-----MADCI 173
Qy 228 KIQMPQYVVHLSNGSKELMEVAD-----NVAR--RVLRHITESLGRVSEDLFALIN 279
Db 174 LVNSQFTAAVFKETFFKLSHIDPVLVPSLNTVSFDSVVPKUL--DDLVPKGGKFFLLSIN 232
Qy 280 SVSRGKGODFLQAFYQALQLOIHBKLVPRHIAVVG---SDVNAQTKEFTQLRDPVVK 336
Db 233 RYERKKULTLALVALQJLRGLTSGDWE--RYVLIYAGGYDBRVLEVHEHGYELKMGVQ 290
Qy 337 NTIHDVHFV---NKTILAVAYTLAIDVLYONSQGRGEGCFRTITIEAMFKLPVIGTAG 393
```

```
Db 291 SBLGQVYVFLRSPDKOKISLHSCCTLVLYTPS---NHRFGIVLEMYWQCPVIAVNSG 347
Qy 394 GTTEIVLDGSTGLHPAGKEGVAFLAKNIIVRLASHAEORVSGEKGYGKEMFEMHHMA 453
Db 348 GPLESIDHSTVGFCLCPDPVHFSRIEKFIRPS---LKATMGLAGARVKEKFSBEAFT 404
Qy 454 ERI 456
Db 405 EQL 407

RESULT 6
US-09-738-626-4738
; Sequence 6, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4738
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4738

Query Match 5.6%; Score 133.5; DB 9; Length 409;
Best Local Similarity 23.7%; Pred. No. 0.00075;
Matches 86; Conservative 52; Mismatches 160; Indels 65; Gaps 13;

Qy 133 NHGVQVLPARGQEARVDIALKADIVILNTAVAGKMLDVLKDH-VFKVL-----PKILMW 185
Db 81 NPAIKTL-STGIRMAEAAANNVDVSHSHWYAGLGHLAALHGIPIHVAHSLERDPK 139
Qy 186 IHMRGHY-----FKVEYVGHLPFVAGAMIDSHTTAEVWNSRSTDRKIQMPQYVV 237
Db 140 REQLGGGYVSWSEKNAMEYADAVIAVSAWKDSILAA-----YPIEDPNRVV 190
Qy 238 HLGNSKEL-----MEVAENVARVLEHRESIGRSDDLFAIINSVSRGGOOLF 291
Db 191 LNGIDTELMOPEPTFDDAEDSVLR-----SLGVDPRPVAFAVGRIKQGVHHL 241
Qy 292 QA---FYQALQLOIHEKLVRIHAAVVGSDVNAQTKEP--TQLRDPVYKNTIHDVHF 345
Db 242 KAAALFDESQVQVLCAGAPDTPFIARFTTALVEELQAKRGKIPWQDMIGKDKIOE---- 297
Qy 346 VNKTILAVAYTLAIDVLYONSQGRGCFRTITIEAMFKLPVIGTAGGTEIVLDGSTG 405
Db 298 -----ILTAADTFVCS--IYPLGIYVLEAMCMTAVASVSGIPEVVVDGTTG 346
Qy 406 LHPAGKEGVAFLAKNIY---RLASHAEORVSGEKGYGKEMFEMHHMAERIAAVLK 461
Db 347 ALVHYDENDVEFFERDIAEAVNHQVADRETAATGLAGREERAINPSWATIAOQTIIDVYK 406
```

Qy 462 DVL 464  
Db 407 SLM 409

RESULT 7  
US-09-738-626-3951

Sequence 3951, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAOKO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 3951  
LENGTH: 418  
TYPE: PRF  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-3951

Query Match 5.4%; Score 128.5; DB 9; Length 418;  
Best Local Similarity 21.1%; Pred. No. 0.0022;  
Matches 82; Conservative 47; Mismatches 151; Indels 109; Gaps 16;  
Qy 64 AAGSSPLGFWRSKLV--LVSHLSLGGPFLMELAFLE-----HVSQVWIT 113  
Db 86 AAFGTGMLSTFRREKTYDILHSHYWSG-----QVGMILDMRIPLIHTAHTLAIVK 139  
Qy 114 NORSQETNDVYSLEHRLMHGVLPARGQEAVIDALKADLVLTINTAVAGKMLDPVLDK 173  
Db 140 NS-YRDSDPPESEARICE-----QQLVD--NADVLAVNTQEE----- 175  
Qy 174 HVPKVLPKILMWHKRGHYFVEVVKHLPFVAGAMIDSHTTAEVWNSRTSDRLKIQMPQ 233  
Db 176 -----MODLMHHY-----DADPDRISSVSP- 195  
Qy 234 TVVHLGNSKELMEVADNVARVLRHRESLGYRSEDLFAIINSVSRKGODLFLQA 293  
Db 196 -----GADVELYSFGNDRATERSRR-----LGRIHTKVVAVGSLQPFKGPVLLKA 244  
Qy 294 FYQALQILQHEKXKVRHIAVVGSDVNAQTKETQLDPPVAKTIDRHVFN-----KTL 350  
Db 245 VAALFDDPPRNLRV-----IICGPGPNATPDT-YHMAEELGVKRIKRLDRPPSE 298  
Qy 351 AVAPYLAIDVLVONSQSGECFGRITTEAMAKLPVLGTAAAGTTEIVLDSTGLLHPA 410  
Db 299 LVAVYRAADIIVAVPSF-----NESFGVAMEAAGTFVIAAVGGLPIAAVBETGLL--- 352  
Qy 411 GKEGVAP--LAKNIYVLAISHAEQVSMGE 437  
Db 353 -VDGSHPAWADALATLDDDETIRRMGE 380

RESULT 8  
US-09-738-626-5896

Sequence 5896, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: TATEISHI, NAOKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
CURRENT FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 5896  
LENGTH: 381  
TYPE: PRF  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-5896

Query Match 5.3%; Score 127; DB 9; Length 381;  
Best Local Similarity 25.2%; Pred. No. 0.0027;  
Matches 55; Conservative 33; Mismatches 82; Indels 48; Gaps 7;  
Qy 263 RESIGVASEDLFAIINSVSRKGODLFLQAFYQALQIHEKXKPRHIAVNGS----- 318  
Db 169 RKKGFTDTTPVIAVNSRLVPRKQDSLIKAMPVY-----AAPPDQILTVSGRGYE 241  
Qy 319 -----DVNAQTFETQLDPPVAKTIDRHVFNKTLAVAPYLAIDVLVONSQSGRG 370  
Db 242 STLRRLATVDSQNVKFLGRLEVDPMINT-----AAADIFAMPARTRG 284  
Qy 371 -----ECFGRITTEAMAKLPVLGTAAAGTTEIVLDSTGLLHPAKGE--VAPLAKNIY 423  
Db 285 GGLDVEGLGIVLELQAGVPIVAGTSGAPETVTP-ATGLV-----VSGSDVXLSELL 339  
Qy 424 RLASHAEQVSMGEKGYGVKEMEHMAERIAVVK 461  
Db 340 ELDDPIRRAAMGAGRAHVAEAWSEWIEGERTLNIILQ 377

RESULT 9  
US-09-815-028-5  
Sequence 5, Application US/09815028  
Patent No. US20020068823A1  
GENERAL INFORMATION:  
APPLICANT: Meyers, Rachel A.  
TITLE OF INVENTION: 33877 AND 47179, NOVEL HUMAN  
FILE REFERENCE: 10448-033001  
CURRENT APPLICATION NUMBER: US/09/815,028  
CURRENT FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: US 60/191,964  
PRIOR FILING DATE: 2000-03-24  
NUMBER OF SEQ ID NOS: 8  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5  
LENGTH: 416  
TYPE: PRF  
ORGANISM: Homo sapiens  
US-09-815-028-5

Query Match	4.9%;	Score 118;	DB 9;	Length 371;
Best Local Similarity	24.2%;	Pred. No. 0.017;		
Matches	50;	Conservative 40;	Mismatches 91;	Indels 26; Gaps 9
Qy	259	REHHSISGVSEDLPAIIINVSRSKQODLFQAFYQALQIQHEKLVKPIRHAVVVS	318	
Db	182	REIYRQNGIKGQNLILVQSGDPFKRVDKRSIEALASPSLSHNTL-----LFPVQG	235	
Qy	319	DVNAQKRFETQLRDPFVVKYKTIHDRVHPNKTALVAVYLAIDLVYQNSGRCGCGSITL	378	
Db	236	D--KPKRFPA-----LAEKLGKSVNHWHPFSGENDVSLIMAAADLILHRY--QEAAGVIL	287	
Qy	379	EAMAFKLPVLGTAAGCTTEIIVDSTG--LIIHPAGKEGVAPLAKNIYRLA-SHAERQVSM	435	
Db	288	EATPAGLPVLTTPAAGVAHYADANCGVLIAPSPSEQL---NEVIRKLTQSPSLRW	343	
Qy	436	GE--KGQYGVKEMEHMHHMARIALV	460	
Db	344	AENNAHYADTDQL--SLPEKAAADI	367	

APPLICANT: Zhang, Zemin  
TITLE OF INVENTION, SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS

Db 435 HVESYL-EAINALTOOTSVLKRS 456

RESULT 12  
US-10-174-590-296  
; Sequence 296, Application US/10174590  
; Publication No. US20030008352A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Chen, Jian  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Pan, Jiege  
; APPLICANT: Smith, Victoria  
; APPLICANT: Matenabe, Colin K.  
; APPLICANT: Wood, William I.  
; APPLICANT: Zhang, Zhenlin

1 TITLE OF INVENTION: ACIDS ENCODING THE SAME  
 2 FILE REFERENCE: P3430R1C42  
 3 CURRENT APPLICATION NUMBER: US/10/174,550  
 4 CURRENT FILING DATE: 2002-06-18  
 5 Prior application removed - See File Wrapper or Patent  
 6 NUMBER OF SEQ ID NOS: 612  
 7 SEQ ID NO: 296  
 8 LENGTH: 323  
 9 TYPE: prt  
 10 ORGANISM: Homo Sapien  
 11 US-10-174-550-296

Query Match 4.6%; Score 111; DB 9; Length 323;  
Best Local Similarity 22.6%; Pred. No. 0.062;  
Matches 67; Conservative 46; Mismatches 143; Indels 40; Gaps 11.

```

QY 181 KILAMW-H-----EMRCHYFKAEYVKHJPRVAGAMJDSHTAEVMSRFSDBLKIQMPC 234
D 39 KILFYCHPFDLLLYTKRBSFKLYRPAIDW-----IEEYTG-----MADOLVNSQFT 87
QY 235 YVHLNLSKEJMEVAED-----NVAR--RVLEHNRRESIGVSEDDLPAIIVNSVNSGG 286
D 88 AAFVKEFTEKSLSHIDPVLVPSLVNTSPDSVBPBK--DVLVPKGFKELLSINREKKN 146
QY 287 ODVFOAFQALQIQIEKELKVPRIHVVWG---SDVNAQYKTEPTQDRPVYKXTHIDRY 343
D 147 LTLAEALVQVRGRLTSPQWE--RVHLIVAGSIDERVLNENYVQELKKNQOOSDQLQY 204
QY 344 HFV--NKTALVAVYALDVLVQNSQGRBECGKITTEMAFKLYVLGTAACTTEYVL 400
D 205 TFLRSQDKQKLSILHSCTVLYTP--NEHGVLPLEAMVMOCCVLAIVNSGGPELSD 261
QY 401 DSGTGLHPAGKGVAPLAKIVRLASHABQVRSVGEKGVGVCEMFMENHMAERI 456
D 262 HSYTGLECDPYNHVSALIEKFIKEPS--LKATVGLAGRAVKEKFSPEAFETOL 314

```

RESULT 13  
US-10-176-758-296  
; Sequence 296, Application US/10176758  
; Publication No. US20030008353A1  
; General INFORMATION

1 APPLICANT: Baker, Kevin P.  
 2 APPLICANT: Chen, Jian  
 3 APPLICANT: Desnoyers, Luc  
 4 APPLICANT: Goddard, Audrey  
 5 APPLICANT: Godowski, Paul J.  
 6 APPLICANT: Gurney, AUSTIN L.  
 7 APPLICANT: Pan, James  
 8 APPLICANT: Smith, Victoria  
 9 APPLICANT: Matanabe, Colin K.  
 10 APPLICANT: Wood, William I.  
 11 APPLICANT: Zhang, Zemin  
 12  
 13 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
 14 FILE REFERENCE: P3430R1C104  
 15 CURRENT APPLICATION NUMBER: US/10/176,758  
 16 CURRENT FILING DATE: 2002-06-21  
 17 Prior Application removed - See File Wrapper or Palm  
 18 NUMBER OF SEQ ID NOS: 612  
 19 SEQ ID NO: 3

```

; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-758-296

```

Query Match 4.6%; Score 11; DB 9; Length 32;  
Best Local Similarity 22.6%; Pred. No. 0.62;  
Matches 67; Conservative 46; Mismatches 143; Indels 40; Gaps 11

QY  
181 KILIMWII-----EMGSHFKVEYKRLPRVAGAMIDSHTAETVWMSRSDSLKIQLPOT 224  
||| ||| : ||| :  
39 KILFCHRPDLITKDSTKLIRAPID-----ILEETYG-----MADDTLIVNSQFT 87

QY 235 YVHHLGNSXMLEMVAEAO-----NVAR--RVLEBHRSLGVRSEDLLFALINVSVRKG 266

Db 88 AAVFKRFTFKSLSHIDPVLVPSLANTSPDSVPEKI--DVLVPKKKFFLLISINRYERKKN 146

QY 287 QDLFLQAFYALQIIOHEKIKVPRIIHAVVVG--SDVAQTKRFTQLDRPVKKNIHDRV 343

Db 147 LTLTLEALVQLRGRILTSQDWE--RVHLILVAGGYDERVLENEHYQLKKNVQGDSDGYV 204

QY 344 HFV---NKTIAVAVYLAIDLVLVNSQGRBECGRITTEMAFKPLPLGTAAGTTEVL 400

Db 205 TPLRSPSDKQKISILHSCTCVLYPS---NEHFGLVPELEANTMOCPLVWNSGGLESID 261

QY 401 DGSSTGLIHPAGKEGVAPLAKNIVLLASHAQVRVNSGKGVYRVEMPMENHHAERI 456

Db 262 HSYTGTGLCEBPVHFAEBALKEKIRPES---LKTMTGLAGRAKRVKEKTSSEATTEQL 314

```

RESULT 14
US-10-175-737-296
? Sequence 296, Application US/10175737
? Publication No. US20030013153A1
? GENERAL INFORMATION:
? APPLICANT: Baker, Kevin P.
? APPLICANT: Chen, Jian
? APPLICANT: Desnoyers, Luc
? APPLICANT: Goddard, Audrey
? APPLICANT: Godowski, Paul J.
? APPLICANT: Gurney, Austin L.
? APPLICANT: Pan, James
? APPLICANT: Smith, Victoria
? APPLICANT: Watanabe, Colin K.
? APPLICANT: Wood, William I.
? APPLICANT: Zhang, Zemin
? TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
? FILE REFERENCE: P3450R1C50
? CURRENT APPLICATION NUMBER: US/10/175,737
? PRIOR APPLICATION REMOVED: 2002-06-19
? NUMBER OF SEQ ID NOS: 612
? SEQ ID NO 296
? LENGTH: 323
? TYPE: PRT
? ORGANISM: Homo Sapien
? US-10-175-737-296

```

Query Match 4.6%; Score 11; DB 9; Length 323;  
 Best Local Similarity 22.6%; Pred. No. 0.062;  
 Matches 67; Conservative 46; Mismatches 143; Indels 40; Gaps 11

QY 181 KILMIIH-----EMGXFVKEVKEHFFVGAADIOSHTAAYVMSRRSRLKIQMPOT 234  
 Db 39 KILFCHPDILLITRKDSFSLKRLVAPIDM-----IEEYTTG-----MADCILVNSQFT 87  
 QY 235 YVHFGNSKELMEVAED-----NVAR--RVLRHIREISGVSEBDLLFALINSVSGMG 286  
 Db 88 AAVFKEFETKLSHIDPDVLVPSLNTVTSFDSVPEKL--DDLVPKCKKFFLLSINRYERKKN 146  
 QY 287 QDLFLQAFQADLQIHLQHEKLVPRHIAVVVG---SDVNAQTRKEQGLDDPVKXTHIDVY 343  
 Db 147 LTLALHELVYDRRLRLSDQME--RVHLYIYAGSIDERVLNENYVQELKVMQOSDGLQIV 204  
 QY 344 HGV---AKTLAAVAPYALADVLVYVNSQSGEGEFGRTITTEMAFKPLVLTGAAGTTEIVL 400  
 Db 205 TILNSFSGKCKSLHSGCTCVLYTPS---NEHFGLVPLEMTVMOCCVAVNSGGPELESID 261  
 QY 401 DQSGTGLHPAGKEGVAPLAKNYVLASHAQKSVSGEKGYGRVKEMEHHMAERI 456  
 Db 262 HSGVGFCEPDDPVHSEALIEKFLREPS--LQATGTGLAGARVKKKFSPEAFETQL 314

## RESULT 15







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using bw model

Run on: June 30, 2003, 16:24:52 ; Search time 43 Seconds  
(without alignments)  
1057,479 Million cell updates/sec

Title: US-09-938-294-45

Perfect score: 2397  
Sequence: 1 MAKTPSPFAVAAGRGKGVH.....ERIAAVLKDVLRKSEHSRS 473

Scoring table: BL0SUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR 73 : \*  
1: p1r1 : \*  
2: p1r2 : \*  
3: p1r3 : \*  
4: p1r4 : \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1478	61.7	402	2	G96784 hypothetical prote
2	1471	61.4	458	2	A86330 hypothetical prote
3	482.5	20.1	670	2	B96564 hypothetical prote
4	207	8.6	382	2	D98289 glycosyltransferas
5	207	8.6	382	2	A82994 glycosyltransferas
6	197	8.2	416	2	G71096 hypothetical prote
7	187.5	7.8	388	2	E72354 probable hexosyltr
8	181.5	7.6	381	2	F71196 probable hexosyltr
9	172.5	7.2	370	2	B82751 lipopolysaccharide
10	166.5	6.9	377	1	A70004 probable hexosyltr
11	166	6.9	382	2	A02858 hypothetical prote
12	161	6.7	405	2	B96817 hypothetical prote
13	156	6.5	416	2	B95920 probable membrane-
14	154.5	6.4	351	1	H64446 probable hexosyltr
15	153	6.4	381	2	C41317 probable lipopolys
16	151.5	6.3	411	2	H69158 LPS biosynthesis R
17	151	6.3	375	2	C83860 hypothetical prote
18	149.5	6.2	393	2	H72352 lipopolysaccharide
19	149	6.2	381	2	A00973 lipopolysaccharide
20	148.5	6.2	378	2	H83019 probable glycosyl
21	147.5	6.2	351	2	H97621 lipopolysaccharide
22	147.5	6.2	351	2	A82844 hypothetical prote
23	147	6.1	333	2	F84166 LPS glycosyltransferas
24	145	6.0	395	2	H84113 hypothetical prote
25	144	6.0	409	2	S76126 hypothetical prote
26	144	6.0	429	2	AC2227 hypothetical prote
27	142	5.9	349	2	A82360 glycosyl transfera
28	142	5.9	329	2	H87379 glycosyltransferas
29	142	5.9	381	2	AH2160

30	141.5	5.9	412	2	T35514 probable glycosyl
31	141	5.9	354	2	A63315 lipopolysaccharide
32	140	5.8	380	2	B97275 glycosyltransferas
33	140	5.8	429	2	A82456 glycosyltransferas
34	138.5	5.8	427	2	B95936 probable glycosylt
35	138	5.8	316	2	F70441 capsular polysacch
36	138	5.8	344	2	C81152 LPS biosynthesis p
37	136	5.7	376	2	A82676 conserved hypothet
38	136	5.7	422	2	AH2352 hypothetical prote
39	135	5.6	385	2	D97911 hypothetical prote
40	134.5	5.6	390	1	A75059 probable hexosyltr
41	133.5	5.6	230	2	G81870 probable glycosyl
42	132.5	5.5	417	2	A82359 hypothetical prote
43	131	5.5	364	2	H75466 lipopolysaccharide
44	131	5.5	379	2	S77338 LPS glycosyltransferas
45	131	5.5	430	2	A12455 glycosyltransferas

## ALIGNMENTS

## RESULT 1

G96784 Hypothetical protein F1B16.5 (imported) - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence\_rev1stion 02-Mar-2001 #ext\_change 31-Mar-2001

C/Accession: G96784

R/Theologian, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.,

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Kuykin, E.; Kim, C.

C.A.; Li, J.H.; Li, T.; Lin, X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, I.

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A6141; MUID:21016719; PMID:11130712

A/Accession: G96784

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-402 <STO>

A/Cross-references: GB:A805173; NID:910120445; PIDN:AA013070.1; GSPDB:GN00141

C/Genetics:

A/Gene: F1B16.5

A/Map position: 1

Query Match 61.7%; Score 1478; DB 2; Length 402;

Best Local Similarity 74.5%; Pred. No. 9,66-102;

Matches 283; Conservative 53; Mismatches 44; Indels 0; Gaps 0;

QY	89	SGGPIILMELAFLLRHVGSOVWITNQSQETNDVYSLHRLAHGVOYLPARGQAND	148
DB	18	SRGPIILMELAFLLRHVGSOVWITNQSQETNDVYSLHRLAHGVOYLPARGQAND	77
QY	149	IALKADLVITAVAGKMLDPVLDKHPVPLPKLIMWHEHMGHFKVEYVKKLPPVAG	208
DB	78	TSKALVITAVAGKMLDPVLDKHPVPLPKLIMWHEHMGHFKVEYVKKLPPVAG	137
QY	209	MDSHTTAAVWNSRTSRLKIQNPQTYVHHGNSKEIMAEENNAKRYLEHRELSG	268
DB	138	MDSHTTAAVWNSRTSRLKIQNPQTYVHHGNSKEIMAEENNAKRYLEHRELSG	197
QY	269	RSEDLFAITNSVSGKQDLFLQAFYQALQIOHEKIKAPRIHAVVGSQVNAQTFFET	328
DB	198	RSEDLFAITNSVSGKQDLFLQAFYQALQIOHEKIKAPRIHAVVGSQVNAQTFFET	257
QY	329	QLRDFFVNTTIDRVHFNKTLAVAPYLAIDVYVNSQSGRCERTITAMAFKLPVL	388
DB	258	ELRPFREKKLENFVHFNKTLAVAPYLAIDVYVNSQSGRCERTITAMAFKLPVL	317
QY	389	GTAAGTTIVDGSGLHPAKGEGVAPLAKNTVRLASHAEORVSMKGYGVKEMFM	448

Db 318 GTAAGTMEIVNGTGLTSLHSAKSGVPLAKNIVKLATQVELRLRMGKNYERVKEMFL 377  
Qy 449 EHHMARIAVLDVLRKSG 468  
|||:||||:|:|:  
Db 378 EHHMSRIASVLRKEVLOHAK 397

## RESULT 2

A66330  
hypothetical protein F69.24 - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Dec-2001  
C/Accession: A66330  
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzilli,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A66141; MUID:21016719; PMID:11130712  
A/Accession: A66330  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-458 <STO>  
A/Cross-references: GB:AE005172; NID:g10086496; PIDN:AMG12556.1; GSPDB:GN00141  
C/Genetic:  
A/Map position: 1

Query Match 61.4%; Score 1471; DB 2; Length 458;  
Best Local Similarity 62.8%; Pred. No. 3.8e-101;  
Matches 304; Conservative 65; Mismatches 73; Indels 42; Gaps 6;

Qy 1 MAKTPSFAVAAGRGVPHNRFTLLLLVAAVASASTAGFLRGALRDPDCCGSG--- 56  
|||:||||:|:|:  
Db 1 MAKPESTSWMAATLQKKEMP-----LMTLLV--LSVTVGMILVRSFPGSSGVSXGCS 50  
|||:||||:|:|:  
Qy 57 -----DPAALNTAVASGS--ELGPMRSKVLVNHSHLSGSPFLIMLALRLHNGSOVY 110  
|||:||||:|:|:  
Db 51 REKEDNSDIKIGSVSSSLNPLLEPMKSKVLVLSHSLSSGGFLMLLAFLLRGESEV 110  
|||:||||:|:|:  
Qy 111 WITNRSQETNDVYSLERHMLNFGVYLPARGQEAVIDAKDLVLTNTAVAGKMLDPV 170  
|||:||||:|:|:  
Db 111 WITNCKVEDEVIKYLEHMKMLDRGVQVYISAKQKADTALKSDLVVLTNTAVAGKMLD 170  
|||:||||:|:|:  
Qy 171 LKDNVPRVLPKILMWIHEMRGKVFVEYVGHLPFVAGAMIDSHSTAAYMNSRPSRLKIQ 230  
|||:||||:|:|:  
Db 171 LKDNVPRVLPKILMWIHEMRGKVFVEYVGHLPFVAGAMIDSHSTAAYMNSRPSRLKIQ 230  
|||:||||:|:|:  
Qy 231 MPQTYVHNLGNSKELMEVADNVARVLRHRESLIGVRSBDLLFALINSVSRGQDLF 290  
|||:||||:|:|:  
Db 231 MPQTYVHNLGNSKELMEVADNVARVLRHRESLIGVRSBDLLFALINSVSRGQDLF 290  
|||:||||:|:|:  
Qy 291 LQAFYQALQLOIH-EKLAKPRIHAVVVGSVVNAQTKFETOLDPVPVKNTHIRVFNKT 349  
|||:||||:|:|:  
Db 270 LRAHESLKVITKTKLEVPFMHNAVVGSDMSQIKFETELRNVQEMKQKIVAFVFNKT 329  
|||:||||:|:|:  
Qy 350 LAVAPYLAIDVLYVNSQSGEGCGRTTEANAFKLPVLGTAAGTTEIVLDGTTGLHP 409  
|||:||||:|:|:  
Db 330 MKVAPYLAIDVLYVNSQSGEGCGRTTEANAFKLPVLGTAAGTTEIVNRRTGLLHN 389  
|||:||||:|:|:  
Qy 410 AGEKGVAPLAKNIVRLASHAEQVSMGEKSGVGVKEMFEMHMAERLAIVLDVLRKSG 469  
|||:||||:|:|:  
Db 390 TGDGVLPPLAKNIVKLATNTVOKRNTYKQGYERVENEMFLHHNSHRIASVLRKEVLOHAKI 449  
|||:||||:|:|:  
Qy 470 HSR 473  
|||:||||:|:|:  
Db 450 HSR 453

## RESULT 3

E96564  
hypothetical protein F6D8.36 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C/Accession: E96564  
R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marzilli,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A66141; MUID:21016719; PMID:11130712  
A/Accession: E96564  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-670 <STO>  
A/Cross-references: GB:AE005173; NID:g5903062; PIDN:AMD55621.1; GSPDB:GN00141  
C/Genetic:  
A/Map position: 1

Query Match 20.1%; Score 482.5; DB 2; Length 670;  
Best Local Similarity 29.2%; Pred. No. 9.7e-28;  
Matches 148; Conservative 77; Mismatches 173; Indels 109; Gaps 13;

Qy 38 STAGFLIR--GALRD-----PCDGRGDPALNTAVASGSPDGPMK---SKVL 80  
|||:||||:|:|:  
Db 193 ATYKGLGPFSGLEDEKVLWSPHRRSGTCDRKS-----FKVLWSRRFV 237  
|||:||||:|:|:  
Qy 81 LVSHSLSGCPFLIMELAFRLRHVGSQVYVWITNRSQETNDVYSLERHMLNFGVYLP 140  
|||:||||:|:|:  
Db 238 LTFELSLMTGALISMELASBLSCATYSAVLSRSG-----GLMELSRRIKIVE 290  
|||:||||:|:|:  
Qy 141 ARQGEAVDIALRADVLVLTNTAVAGKMLDPVLPKILMWIHEMRGKVFYK----- 195  
|||:||||:|:|:  
Db 291 DKGLSFTKPMKADLIAGSVCTSWIDQYMNH-PAGGSQIAMVMEKREYDPRAKV 349  
|||:||||:|:|:  
Qy 196 VEYVKLPFVAGAMIDSHSTAAYMNS-RTSDRLKIQMPQTYVHNLGNSKELMEVADNV 249  
|||:||||:|:|:  
Db 350 LDRYKMLIFL-----SSQSQRLWCEEHKIKR--SQVYIPLSVNDELAPFAGIPSS 402  
|||:||||:|:|:  
Qy 250 -----EDNVARVLRHRESLIGVRSBDLLFALINSVSRGQDLFQAFYQALQ- 299  
|||:||||:|:|:  
Db 403 LNTPTLSPEKMKVRKQILRESVTELTGTDSDMLWSLISINFTGQDLLESLALSLP 462  
|||:||||:|:|:  
Qy 300 -----LIOHEKLVKPRIHAV-----VGS 318  
|||:||||:|:|:  
Db 463 RQGESQRNHKGIIRKREKVLSSKRLRSGSRQMSVSLTLNGLRREKQELKVLGVS 522  
|||:||||:|:|:  
Qy 319 DVNAQKPEFTQDRPVVKNTHIRVFNKTLAAPYLAIDVLYVNSQSGEGCGRTT 378  
|||:||||:|:|:  
Db 523 KSKNGVYKEMLSFSLNSGVLSSVMTPTATTRVSLYSADVYVNSQSGEGCGRTT 582  
|||:||||:|:|:  
Qy 379 EANAFLPVLGTAAGTTEIVLDGTTGLHPAGKEGVAPLAKNIVRLASHAEQVSMGEK 438  
|||:||||:|:|:  
Db 583 EANAFLPVLGTAAGTTEIVLDGTTGLHPAGKEGVAPLAKNIVRLASHAEQVSMGEK 438  
|||:||||:|:|:  
Qy 439 GYGRVKEPMFEMHMAERLAIVLDVLRKSG 465  
|||:||||:|:|:  
Db 643 GRGVNEMKMKQHYKRPVLDVLRKSG 669  
|||:||||:|:|:

## RESULT 4

D98289  
hypothetical protein AGR L2541 [imported] - Agrobacterium tumefaciens (strain C58, Cerec  
C/Species: Agrobacterium tumefaciens  
C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C/Accession: D98289

R.Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirolo, B.; Goldman, A.; Liu, F.; Molism, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens*  
A/Reference number: A97359; PMID:11743194  
A/Accession: D98289  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-382 <KUR>  
A/Cross-references: GB:AE007870; PIDN:AAK69838.1; PID:g15159774; GSPDB:GN00170  
C/Genetics:  
A/Gene: AGR\_L\_2541  
A/Map position: linear chromosome

Query Match 8.6%; Score 207; DB 2; Length 382;  
Best Local Similarity 27.5%; Pred. No. 1,1e-07;  
Matches 68; Conservative 46; Mismatches 109; Indels 24; Gaps 7;

QY 220 NSRTSDRLKI-----QMPQTVVHLGNSKELMEVADNVARVLRHIRESLGVRSDDLIF 275  
DB 154 NSEETGAPFTEHGEBAUKRIVYNGFDPKAKLHDAGMAR-----LRAELGLGPQPLV- 207  
QY 276 AINSVSRGKGQDLFLQAFYQALQIHEKLVPRHIAVVVSGDVNAQTKETQLRDYV 335  
DB 208 GLFRLSEWKGQHVFLDAL-AAWEGVQ-----AVIVGALFGQEAVERARIREQAS 256  
QY 336 KNTIHRYHFVVKTLAAVPLAIDVLVONSQSGECGCRITTEMAFPLVLTAGGT 395  
DB 257 RLGLDRVAFVGFNSDPELMASMDVAHTSI-VAEPGRVVEAMCGRPVAVTRGGV 315  
QY 396 TEIVLDSTGLHPACKEGVAPLAKNIVLASHAEORVSMGEKGYGVKEMFEMHMAER 455  
DB 316 TEIIRDEGTGLVPPG--DASALAAAGTILSDPALAQLRGSGREDVSDRFSLQETCRS 373  
QY 456 IAAVLKD 462  
DB 374 VSAALTE 380

## RESULT 5

AF2994  
glycosyltransferase Atu3560 [imported] - *Agrobacterium tumefaciens* (strain C58, Dupont)  
C/Species: *Agrobacterium tumefaciens*  
C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
R/Accession: AF2994  
C/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, L.; erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCell  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A/Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.  
A/Reference number: AB2577; PMID:11743193  
A/Accession: AF2994  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-382 <KUR>  
A/Cross-references: GB:AE008689; PIDN:AAI44372.1; PID:g17741967; GSPDB:GN00187  
A/Experimental source: strain C58 (Dupont)  
C/Genetics:  
A/Gene: Atu3560  
A/Map position: linear chromosome

Query Match 8.6%; Score 207; DB 2; Length 382;  
Best Local Similarity 27.5%; Pred. No. 1,1e-07;  
Matches 68; Conservative 46; Mismatches 109; Indels 24; Gaps 7;

QY 220 NSRTSDRLKI-----QMPQTVVHLGNSKELMEVADNVARVLRHIRESLGVRSDDLIF 275  
DB 154 NSEETGAPFTEHGEBAUKRIVYNGFDPKAKLHDAGMAR-----LRAELGLGPQPLV- 207  
QY 276 AINSVSRGKGQDLFLQAFYQALQIHEKLVPRHIAVVVSGDVNAQTKETQLRDYV 335  
DB 208 GLFRLSEWKGQHVFLDAL-AAWEGVQ-----AVIVGALFGQEAVERARIREQAS 256

DB 208 GLFRLSEWKGQHVFLDAL-AAWEGVQ-----AVIVGALFGQEAVERARIREQAS 256  
QY 336 KNTIHRYHFVVKTLAAVPLAIDVLVONSQSGECGCRITTEMAFPLVLTAGGT 395  
DB 257 RLGLDRVAFVGFNSDPELMASMDVAHTSI-VAEPGRVVEAMCGRPVAVTRGGV 315  
QY 396 TEIVLDSTGLHPACKEGVAPLAKNIVLASHAEORVSMGEKGYGVKEMFEMHMAER 455  
DB 316 TEIIRDEGTGLVPPG--DASALAAAGTILSDPALAQLRGSGREDVSDRFSLQETCRS 373  
QY 456 IAAVLKD 462  
DB 374 VSAALTE 380

## RESULT 6

G71096  
hypothetical protein PH1035 - *Pyrococcus horikoshii*  
C/Species: *Pyrococcus horikoshii*  
C/Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 05-Nov-1999  
R/Accession: G71096  
R/Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekine  
M.; Onitoku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,  
DNA Res. 5, 55-76, 1998  
A/Title: Complete sequence and gene organization of the genome of a hyper-thermophilic an  
A/Reference number: A71000; MUID:98344137; PMID:9679194  
A/Accession: G71096  
A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-416 <KAW>  
A/Cross-references: GB:AP000004; NID:g3236131; PIDN:BAH30133.1; PID:d1031076; PID:g32574;  
A/Experimental source: strain OT3  
A/Note: this accession replaces an interim accession for a sequence replaced by Genbank  
C/Genetics:  
A/Gene: PH1035

Query Match 8.2%; Score 197; DB 2; Length 416;  
Best Local Similarity 25.5%; Pred. No. 6,7e-07;  
Matches 104; Conservative 56; Mismatches 174; Indels 74; Gaps 18;

QY 87 SLGGPFL-LMEALFLRHVGSQVVTINORSQETNDVYSLERHMLNHGQVLPARGOE 145  
DB 50 SPFGGVAELHSLVPLRLRSIGIEARWVIEGPFEPFVTKTHNL-----QSGNE 99  
QY 146 AVIDALKAIDLILN-TAVAGKMLDPLKDHVPKVLKILMIWHENRGHYFVYVGHLP 204  
DB 100 SLCTPENKELIYLVNENSKPILDSFDV-----LVHDPQPALIEFPEKKSFW 150  
QY 205 VAGAMID-SHTTAAYN-----SRTSDRLKIQMPQTVVHLGNSKEL-----EVAEDN 252  
DB 151 IMRCHIDISSPREFEFELRRFVEKDYIFHLPEYVQPELRNRAVIMPPSIDPUSKX 210  
QY 253 VARR--VLRHIRESLGVRSDDLFLAIIINSVR--GKGQDLFLQAFYQALQIHEKX 306  
DB 211 VELKQTEILR--ILRPDVDPK--PIITQVSRFDMKG-----IFDVIEIYRKYXE 258  
QY 307 KVRPIHAAVVG--SVNAQTKETQLR--DFVVK--NTIHDRHVFNKTAAVAP 355  
DB 259 KIGVQVLILGVMAHDPDEGMVIFEXTLRKIGEDYDVKVLITNLIGHARVNA-----AF 312  
QY 356 IAAIDVLVONSQSGECGCRITTEMAFPLVLTAGGTTEIVLDGSTGLHPACKEGV 415  
DB 313 QRSVDIILGMSIRBG--FGLTVEAMWKGKPVIGRAVGIGIKQIYDGETGLVDANEAV 370  
QY 416 APLAKNIVLASHAEORVSMGEKGYGVKEMFEMHMAERIAAVLKDV 463  
DB 371 ----EKVLYLHGPVSKEMGAKARVKNPITTKMERKYLIDLSL 414

## RESULT 7

E72354  
probable hexosyltransferase (EC 2.4.1.-) TM0622 - *Thermotoga maritima* (strain MSB8)  
C/Species: *Thermotoga maritima*

[illegible]

Db DERFFAPRODKETIKKFGLEGNVLVY--VSRNSYKKGSHVLNHF-----SKTE 229  
QY 310 RHAHVVSDD-----VNAQTKFETQLRDFVFNKQTIHDFVFNKTLAVALPYLAIDVLYVQ 364  
Db 230 DATLVMVNGENMLPFLKAQTKF-----LGIEKNVCFVMGVVPDDILPEVFRADDFVLPS 283  
QY 365 NSGORECEGRTTIEMAFKLPVLGFLNAGGTETIYLDGSGLLHPKPGKGVAPLAKNIVR 424  
Db 284 IS--SEAGVGIYILEMANSGVPIIADVDGIGPEVYKENSGLLVPPGNE--LKLREALEK 338  
QY 425 LASHAEQVRSVMGEKGYGRVKEEMFEHMAERLAVALKDVRK 466  
Db 339 LKNEELRKVYGNNGRSVBEKYSNMKIYVKIERIYNELDGE 380  
RESULT 9  
E82751  
lipopolysaccharide biosynthesis protein Xf0879 (imported) - *Xylella fastidiosa* (strain 9606)  
C:Species: *Xylella fastidiosa*  
C:date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: E82751  
R:anonymouse, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing  
Nature 406, 151-157, 2000  
A>Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
A:Reference number: A82515; MUID:20365717; PMID:10910347  
A:Note: for a complete list of authors see reference number A59328 below  
A:Accession: E82751  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-370 <STM>

A/Citomeb-references: GB:AE003927; GB:AE003649; NID:9105783; PMID:AA083689.1; GSPDB:GN0014  
A/Experimental source: strain 945C  
R/Simpson, A.J.G.; Reinach, P.C.; Artuda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Al  
Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrex, H  
as-Neto, E.; Docena, C.; El-Dorriy, H.; Facinanti, A.P.; Ferreira, A.J.S.  
submitted to Genbank, June 2000  
A/Authors: Ferreira, V.C.A.; Ferriz, J.A.; Fraga, J.S.; Franco, M.C.; Fromme  
J.D.; Jungueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.F.; Kurama, E.E.; Laigret  
Chado, M.A.; Madrita, A.M.B.N.; Madeira, H.M.F.; Marinho, C.L.; Marques, M.V.; Martins, E  
A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.  
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasaki  
A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira  
W., Tsunako, M.H.; Valada, H.; Van Sluys, M.A.; Verjovsky-Almeida, S.; Vettore, A.L.; Ze  
A/Reference number: A59388  
A/Contents: annotation  
C/Genetics:  
A/Gene: XF0879







```
Db      249 KCSFLVPS--RSEGFQVAVBGMCSKVATATRVGGLGEIVDGYNGLL--AEKNPNPND 304
Qy      418 LANNIVRLASHAEQVRVSMGEKG 439
Db      305 LKEKILIELINNEELRKTILGENG 326
```

## RESULT 15

```
C41317
probable lipopolysaccharide N-acetylglucosaminyltransferase (EC 2.4.1.56) - Salmonella
C:Species: Salmonella typhimurium
C:Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 08-Oct-1999
C:Accession: C41317
R:MacLachlan, P.R.; Kadam, S.K.; Sanderson, K.E.
J. Bacteriol. 173, 7151-7163, 1991
A:Title: Cloning, characterization, and DNA sequence of the rfaK region for lipopolysac
A:Reference number: A41317; MUID:92041612; PMID:1657881
A:Accession: C41317
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-381 <MAC>
A:Cross-references: GB:M73826; NID:g154328; PID:AAA27207.1; PID:g454851
C:Keywords: glycosyltransferase; hexosyltransferase
```

Query Match 6.4%; Score 153; DB 2; Length 381;

Best Local Similarity 23.4%; Pred. No. 0.0011;

Matches 74; Conservative 40; Mismatches 110; Indels 92; Gaps 11;

```
Qy      168 DPLVAKDHPVRLPKILMWIHEMRGHYFKVEYVKHLPEVAGAMIDSHTEYVNSRTSDRL 227
Db      133 EPELPNDAKII-----VPSOFLKAFYEERLPAAAVSIVPNGCAETVKRNPQDNL 183
Qy      228 KIQMPQTVVYHLGNSKELMEVADNVARRVLRHIRESLGVRSDDLPAIINSVSRGQG 287
Db      184 RQG-----LNIADATV-----LLVA--GRISPDGI 208
Qy      288 DPLQAPYALQILQHEKLVPRILHAVVVGSDVNAQTKFETQLRDFVVKNTIHDRHFPN 347
Db      209 LLLLOAFKQLRTRSNIKL-----VVVG-DPVASRKGE-----KAEYOK 246
Qy      348 KTLAVVPVLA-----IDVLVNSQSGEGCFGRITTEMAFKLPVL 388
Db      247 KVLDAKKEIGTCIMAGQSPQMHNFYHIALVIVPSQVE-EAFQVAVEMAAAGKAVL 305
Qy      389 GTAAGTTEIVDSTG--LIHPAGEGVAPLAKNIVRLASHAEQVRVSMGEKGYGVKEM 446
Db      306 ASKKGGISSEFVLDGITGYHLAEPWSDSI1---NDINRALADKERHQIAEKXSLVFSK 361
Qy      447 FMEHMAERIAAVLKD 462
Db      362 YSMENVAQRFEEQMKN 377
```

Search completed: June 30, 2003, 16:30:26  
Job time : 45 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2003, 16:26:57 ; Search time 22 Seconds  
(without alignments)  
891.741 Million cell updates/sec

Title: US-09-938-294-45

Perfect score: 2397

Sequence: 1 MAKTPSPVAVAAGRGPHV.....ERTAAVLDVAKSGSHRS 473

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	185.5	7.7	352	1	LPCC_RHILV
2	180.5	7.5	351	1	LPSE_RHIME
3	166.5	6.9	377	1	CTSA_BACSU
4	153	6.4	381	1	RPAK_SALTY
5	134.5	5.6	380	1	CAPM_STAYU
6	119.5	5.0	428	1	Y486_MYCLE
7	119.5	5.0	480	1	Y486_MYCTU
8	118	4.9	374	1	RPAK_ECOLI
9	118	4.9	578	1	VIRG_SALTY
10	117	4.9	608	1	UQST_MANS
11	112.5	4.7	2564	1	SPOQ_MAIZE
12	108.5	4.5	605	1	UQST_MAIZE
13	107	4.5	340	1	LPSE_RHIME
14	107	4.5	603	1	UQST_HORVU
15	107	4.5	862	1	HEIX_METH
16	106	4.4	358	1	YC07_GLEPN
17	105.5	4.4	508	1	K020_HUMAN
18	105.5	4.4	808	1	SEA2_MYCTU
19	104	4.3	778	1	SEA2_MYCLE
20	104	4.3	805	1	SUSY_PHAU
21	103.5	4.3	608	1	UQST_SORBI
22	103.5	4.3	807	1	SUSI_HORVU
23	102	4.3	353	1	Y938_HAEIN
24	101.5	4.2	536	1	TCPD_FUGRU
25	101.5	4.2	607	1	UQST_SOLTU
26	100	4.2	615	1	UQST_WHEAT
27	100	4.2	801	1	SUS2_DAUCA
28	100	4.2	1158	1	R114_HUMAN
29	99.5	4.2	608	1	UGST_IPOBA
30	99	4.1	804	1	SUS2_ARATH
31	98.5	4.1	1056	1	SPS_SPTOL
32	98.5	4.1	805	1	SUS2_SOLTU
33	98	4.1	377	1	YPJH_BACSU

34	98	4.1	512	1	MCPD_ENTAE
35	96.5	4.0	438	1	GLGA_THECA
36	96.5	4.0	409	1	C71Q_ARATH
37	96.5	4.0	609	1	UGST_ORYGL
38	96.5	4.0	609	1	UGST_ORYSA
39	96.5	4.0	805	1	TP2_AQUAE
40	96.5	4.0	926	1	POOL_HAEIN
41	96	4.0	744	1	RELA_ECOLI
42	96	4.0	778	1	TRKA_CHICK
43	96	4.0	808	1	SUSI_ORYSA
44	96	4.0	1057	1	SPSL_CITUN
45	95.5	4.0	643	1	HS70_CIAHE

## ALIGNMENTS

RESULT 1	ID	LPCC_RHILV	STANDARD	PRT	352 AA
AC	068547				
DT	30-MAY-2000	(Rel. 39, Created)			
DT	30-MAY-2000	(Rel. 39, Last sequence update)			
DT	30-MAY-2000	(Rel. 39, Last annotation update)			
DE	Lipopolysaccharide core biosynthesis mannose 6-phosphate transferase (EC 2.4.1.22)				
DE	LPCC				
GN	Rhizobium leguminosarum (biovar viciae)				
OS	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;				
OC	Rhizobiaceae; Rhizobium				
OX	NCBI_TaxID=387				
RM	(1)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=385				
RX	MEDLINE=98434549; PubMed=9756877				
RA	Kadmas J.L., Allaway D., Studholme R.E., Sullivan J.T., Ronson C.W., Poole P.S., Raetz C.R.H.				
RT	"Cloning and overexpression of glycosyltransferases that generate the lipopolysaccharide core of Rhizobium leguminosarum."				
RU	J. Biol. Chem. 273:26432-26440(1998).				
CC	-1- FUNCTION: ACTS AT TRANSFER OF MANNOSE GROUP TO A 3-DEOXY-D-MONO				
CC	OCTULONIC ACID (KDO) VIA AN ALPHA-1,5 LINKAGE.				
CC	-1- PATHWAY: Lipopolysaccharide core biosynthesis.				
CC	-1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).				
CC	EMBL: AF050103; AAC0515.1; -				
DR	InterPro: IPR001296; Glycosyltransferase				
DR	Pfam: PF00534; Glycosyltransferase				
KW	Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase.				
SO	SEQUENCE 352 AA; 38640 MW; 8P3F5CE20C6CB08B CRC64;				
QY	Query Match	7.7%	Score 185.5; DB 1; Length 352;		
QY	Best Local Similarity	7.7%	Pred. No. 3.1e-07;		
QY	Matches	94; Conservative	45; Mismatches 135; Indels 103; Gaps 18;		
QY	136	VOUVPAGQANDIALKADVILNTAVAGKMDPVYKDHVYKULPKLW	-----	184	
QY	26	VQVLPCCIRGLKIXIA	-----	TLGAGLEBEDLPKIKRQDGLGMRPPA	66
QY	185	-----HIHMRGHYFKV-----EYKHLP-----FVAGAMIDSHTEYWNRS-----T	223		
QY	67	RRRRRVV-HARRNNEVAVGILRLTRMPLKLLFTSAQ-RBHATATKULIRMDAVIAT	124		
QY	224	SDRLK-IQNPQVYVHNGSKELM-----EYADNVAKRVLRHREISLGVRESDLLFAIT	278		

Db 125 SDRSGSELEVPHTVIGH-GVDLALFHPPEAABDGA-----ATGLPGRHLV-GCF 172

Qy 279 NSVSRGGDLPLOAFYQALQLOHEKLYKPIHVVGVSDVNAQTKFTOLRDVYKNT 338

Db 173 GRVNHQKGTDLVRMIMELP--QHTEWT-----AVSGRTAEHVAFDKADVAAG 225

Qy 339 IHDRVHFVNK-----TLAVAPYLAIDVLVONSQGRGECGRTITTEMAAFKLPV 387

Db 226 LSDRIIFLGEVPDIKIMYRRLTLVAVP-----SRNEGGLTPLEMAASRAV 272

Qy 368 LGTAAGGTETVLDSTGTLHPAGKEGVAPLAKNIVRLASHAEORVSKGKGVKXEMF 447

Db 273 VASDAGKAYAEIYVGTGTSV--VAASDGEA-LTRALAPIADPALAAHGENMLRVRANF 330

Qy 448 MEHMAERLAAVLKQVL 464

Db 331 ALEREAIGAVYNSIL 347

## RESULT 2

LPSS RHIME STANDARD; PRT; 351 AA.

AC QGRSN2;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Lipopolysaccharide core biosynthesis mannosyltransferase lpsb (EC 2.-.-.-)

GN LPSS OR R01572 OR SMC01219.

OS Rhizobium meliloti (Sinorhizobium meliloti).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Sinorhizobium.

OX NCBI\_TaxID=3382;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=RCR2011 / S047;

RX MEDLINE=21101807; PubMed=1157937;

RA Lagaree A., Hozbor D.F., Niehaus K., Pich Otero A.J.L., Lorenzen J.,

RA Arnold W., Puhler A.;

RT "Genetic characterization of a Sinorhizobium meliloti chromosomal

RT region involved in lipopolysaccharide biosynthesis.";

RL J. Bacteriol. 183:1248-1258 (2001).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=1021;

RX MEDLINE=21396507; PubMed=11481430;

RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Barut J.,

RA Boistard F., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,

RA Godtke T., Goffeau A., Kahn D., Kiss E., Lelaude V., Masny D.,

RA Pohl T., Portetelle D., Puhler A., Purnelle B., Ramsparger U.,

RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;

RT "Analysis of the chromosome sequence of the legume symbiont

RT Sinorhizobium meliloti strain 1021.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882 (2001).

RC FUNCTION: ACTS AT TRANSFER OF MANNOSE GROUP TO A.3-DEOXY-D-MONO

CC OCTULONIC ACID (KDO) VIA AN ALPHA-1,5 LINKAGE (B7 SIMILARITY).

CC -1- PATHWAY: Lipopolysaccharide core biosynthesis.

CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF193023; AAF06008.1; -

DR EMBL; AL591787; CAC6451.1; -

DR InterPro; IPR001296; Glycosyltransf.1

DR Pfam; PF00534; Glycosyltransf.1; 1\_1

KM Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase;

KM Complete proteome.

SQ SEQUENCE 351 AA; 38662 MW; 98436AB9DB82377A CRC64;

Query Match 7.5%; Score 180.5; DB 1; Length 351;  
Best Local Similarity 24.3%; Pred. No. 7,7e-07;  
Matches 84; Conservative 44; Mismatches 133; Indels 85; Gaps 14;

Qy 170 VLKDHVXKLPKT-----LW-----WIENRGHYKRYEVYKHLPEV----- 205

Db 41 VLDPGLPXSLSVFRDLHLMWRPEGRPCRW-HARR-----VEMFALILDLR 92

Qy 206 -----AGAMIDSHHTAEYVNS-----RTSDRLK--IQMPOTVYVHAGNSKEIM 246

Db 93 MKRLVTFSSAGORRHGT---MSKFLIRMDAVIATSGRTAAVLDPVNTVILHGDTRKQ 149

Qy 247 EVADENVARVLRHIESLGYRSEDLFAIINSVSRGQDLFLQAFYQALQLOHEKLY 306

Db 150 PPDCKTEAKK-----ALGLDPAPKPFVCGCFGRVRRHOGTDLFDSMIALLLPC----- 195

Qy 307 KVPRIHAVVGVSDVNAQTKFTOLRDVFNKNTIHDRVHFVNKTLAVAPYLAIDVLVONS 366

Db 196 -RDPWGAIVAGRATGPHLAFESSELKERAKGLADRLTFVGHTNI PDWYRADLFLV--A 252

Qy 367 QGRGECGRTITTEMAAFKLPVLTAAAGTTETVLDGS--TGLHPACK-----EGVAPLA 419

Db 253 PQWEGFGLTPLEMAATGVAVATDVGAFFSELVYGSSEETGLIIAADLKKAVDAAAFM 312

Qy 420 KNIVRLASHAEORVSKGKGVKXEMFMEHMAERLAAVLKQVL 465

Db 313 DDPRIAA-----AASANGLARSTSKNFAIEQEARMAAVVESLWR 351

## RESULT 3

CTSA BACSU STANDARD; PRT; 377 AA.

AC P46915;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Spore coat protein SA.

GN COTS.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI\_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168 / 60015;

RX MEDLINE=95400496; PubMed=7545510;

RA Abe A., Koide H., Kohno T., Matabe K.;

RT "A Bacillus subtilis spore coat polypeptide gene, cots."

RL Microbiology 141:1433-1442 (1995).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98048467; PubMed=9387221;

RA Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;

RT "Sequencing and functional annotation of the Bacillus subtilis genes

RT in the 200 Kb rrmh-dnaB region.";

RL Microbiology 143:3431-3441 (1997).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=98044033; PubMed=9384377;

RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

RA Azevedo V., Bertorello M.G., Bessieres P., Boletich A., Borcher S.,

RA Borriess R., Bourret L., Brans A., Braun M., Brignell S.C., Bron S.,

RA Choi S.K., Brusch C.V., Caldwell B., Capuano V., Carter N.W.,

RA Demitroff F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,

RA Britan K.D., Erttington J., Fabret C., Ferrari E., Fougier D.,

RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizty E.J., Grandi G.,

RA Ghin S.Y., Glaeser P., Goffeau A., Golightly E.J., Grandi G.,

RA Guisepi G., Guy B.U., Haga K., Haeck J., Harwood C.R., Henaut A.,

RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones D.,

RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,  
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,  
 RA Kurita K., Lapidus A., Lardinois S., Lamber J., Lazarevic V.,  
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,  
 RA Medina N., Melhado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara K., Onoeda B., Park S.H.,  
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
 RA Pressac E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
 RA Sekiguchi J., Sekowaka A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takamori K.,  
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsitz P., Tognoni A.,  
 RA Tostato V., Uchiyama S., Vandenhof M., Vannier F., Vassarotti A.,  
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzneger T.,  
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,  
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,  
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*  
 RT *subtilis*.";  
 RL Nature 390:249-256 (1997).  
 RN [4]  
 RP CHARACTERIZATION.  
 RX MEDLINE=9251148; PubMed=10234840;  
 RA Takamatsu H., Kodama T., Watabe K.;  
 RT "Assembly of the *CotSA* coat protein into spores requires *Cots* in  
 RT *Bacillus subtilis*.";  
 RL FEMS Microbiol. Lett. 174:201-206 (1999).  
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.  
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: D31847; BAA06633.1; -;  
 DR EMBL: AF008220; AAC00219.1; -;  
 DR EMBL: Z99119; CAB5069.1; -;  
 DR Subtilase; BG11381; *CotSA*.  
 DR InterPro: IPR001296; Glycosyltransf.1.  
 DR Pfam: PF00534; Glycosyltransf.1; 1.  
 KW Spontaneous; Transferase; Glycosyltransferase; Complete proteome.  
 SQ SEQUENCE 377 AA; 42912 MW; 1F978E1B79F5E660 CRC64;  
 CC -----  
 Query Match 6.9%; Score 166.5; DB 1; Length 377;  
 Best Local Similarity 24.6%; Pred. No. 11e-05; Indels 49; Gaps 18;  
 Matches 85; Conservative 58; Mismatches 153;  
 QY 134 HGVOVLARGQAVDIALKA--DIVLINTVAGKMLDPVLKDHPVKVLPKILMWHENR 190  
 DB 58 HYVHLEDEYEAVGAELKKSFPDLY-HVGNRPSW-PLTKKQAPDANV-FILSVANEK 112  
 QY 191 GHYFKYVYKHLPPVAGAMIDS---HTLAEMNSRSDSLKIQMOTVYVHLSKSELM 246  
 DB 113 FAYDKISQAE---GEICIDSVAGQVTVSDITGOTTSPFANSKIKTVYSGVDIKY 167  
 QY 247 EVAEDVVARVLRHRESIGVSESDLLFALINSVSGKQDPLQAFYQALQLOHEK 306  
 DB 168 HPMWTNCGRA-REEMSESLGLGKKIVL-FVGRLSKVKPHILLQALPDILE--EH--- 220  
 QY 307 KVRRIHAVVGSVDVAQTKF--ETQLRDPVAVNTI-----DRYHPN--KLVAAPY 355  
 DB 221 --EDVMVFTS-----KPFQNDENNTV--KHLITLGMQMDVHTFQVYKQDIPRL 270  
 QY 356 LAAILDVVNSQGECEFGRTTEAMAFKLPVLTAAAGTTEIVLDGSTG-LHPRAKQG 414  
 DB 271 YTVSDVFCSSQMG-EPLARVHYEAMAGLPITISNRGQNPVEYIEGKKGVIHDF--EN 327  
 QY 415 VAPLAKNIVRLASHAEQVSWGKGGVGVKEMFENHMAEIRIAY 459  
 DB 328 PKQYARINDLSSSEKREKRGKYSRREASFGMGVAAENLSTV 372

RESULT 4  
 ID\_RPKX\_SALTY STANDARD; PRT: 381 AA.  
 AC P26470;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Lipopolysaccharide 1,2-N-acetylglucosaminetransferase (EC 2.4.1.56).  
 RPKX OR MARK OR STM3714.  
 GN *Salmonella typhimurium*.  
 OS *Salmonella typhimurium*.  
 OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC *Salmonella*.  
 OK NCBI\_TaxId=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2.  
 RX MEDLINE=92041612; PubMed=1657881;  
 RA MacLachlan P.R., Kadam S.K., Sanderson K.E.;  
 RT "Cloning, characterization, and DNA sequence of the *rfaK* region for  
 RT *Lipopolysaccharide* synthesis in *Salmonella typhimurium* LT2.";  
 RL J. Bacteriol. 173:7151-7163 (1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McCelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of *Salmonella enterica* serovar *Typhimurium*  
 RT LT2.";  
 RL Nature 413:852-856 (2001).  
 CC -1- FUNCTION: ADDS THE TERMINAL N-ACETYL-D-GLUCOSAMINE GROUP ON THE  
 CC GLUCOSE(II) GROUP OF LPS.  
 CC -1- CATALYTIC ACTIVITY: UDP-N-acetyl-D-glucosamine +  
 CC lipopolysaccharide = UDP + N-acetyl-D-glucosaminyl-  
 CC lipopolysaccharide.  
 CC -1- PATHWAY: lipopolysaccharide core biosynthesis.  
 CC -1- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE.  
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.  
 CC -1- SIMILARITY: SOME, WITH B.SUBTILIS YTXN.  
 CC -1- SIMILARITY: SHOWS VERY LITTLE SIMILARITY TO E.COLI RPAK.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL: M73826; AAA27207.1; -;  
 DR EMBL: AB008872; AAL22573.1; -;  
 DR PIR: C41317; C41317.  
 DR StyGne; SG10339; *rfaK*.  
 DR InterPro: IPR001296; Glycosyltransf.1.  
 DR Pfam: PF00534; Glycosyltransf.1; 1.  
 KW Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase;  
 KW Complete proteome.  
 SQ SEQUENCE 381 AA; 43152 MW; F60F37FF175372C6 CRC64;  
 CC -----  
 Query Match 6.4%; Score 153; DB 1; Length 381;  
 Best Local Similarity 23.4%; Pred. No. 0.00012; Indels 92; Gaps 11;  
 Matches 74; Conservative 40; Mismatches 110;  
 QY 168 DPLVDKHPVKVLPKILMWHENRGVYFVEVYKLPVAGAMIDSHTLAEMNSRSDRL 227  
 DB 133 EEPDLPNDAKIL-----VPSQFLKAFYERLRPAAVSIVPGSCATRYKRPDVL 183  
 QY 228 KIQMOTVYVHLSKSELMVAVEDVVARVLRHRESIGVSESDLLFALINSVSGKQ 287

```

Db      184 RQO-----LNIADPATV-----LLYA--GRISPDGI 208
Qy      288 DLFQAFYQALQLOIHEKXKVPRIHAWVSGDVNAQTKFETQADPFVVKNTIHDRVHN 347
Db      209 LILLQARQKQRTLRSHIKL-----VVQ-DYASRKE-----FAEYQK 246
Qy      348 KTLAVAPYLA-----IDLVNQSQGECEGFRITTEAMAFKLPVL 388
Db      247 KYLDAAEIKGTDCIMAGSQSPDMHNFYHADIIVPSQVE-EAFPCVAVAEAMAAAGKAVL 305
Qy      389 GTAAGTTEIYLDGSTG--LHPAGKEGVAFLAKNIVRLASHAQRVSMGEGYGRVEM 446
Db      306 ASKKGISSEFVLDGIGYHIAEPMSDSII-----NDIRALADERIQIEMKAKSLVFSK 361
Qy      447 FMEHHMAERIAVLKQ 462
Db      362 YSMENVAQREEQMK 377

```

## RESULT 5

```

CAPM_STRAU ID CAPM_STRAU STANDARD; PRT; 380 AA.
AC P39862;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Capsular polysaccharide biosynthesis glycosyl transferase capM
   (EC 2.-.-.-).
GN CAPM.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M;
RX MEDLINE=95050273; PubMed=7961465;
RA Lin W.S., Cunnane T., Lee C.Y.;
RT "Sequence analysis and molecular characterization of genes required
   for the biosynthesis of type 1 capsular polysaccharide in
   Staphylococcus aureus.";
RL J. Bacteriol. 176:7005-7016(1994).
CC -1- FUNCTION: REQUIRED FOR THE BIOSYNTHESIS OF TYPE 1 CAPSULAR
   POLYSACCHARIDE.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC EMBL; U10927; AAA64652.1; -.
DR InterPro: IPR001296; Glycos transf_1.
DR Pfam: PF00534; Glycos transf_1.1.
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 380 AA; 45166 MW; 0264E8376CD1A9EF CRC64;

```

```

Query Match 5.6%; Score 134.5; DB 1; Length 380;
Best Local Similarity 25.3%; Pred. No. 0.0035;
Matches 46; Conservative 38; Mismatches 69; Indels 29; Gaps 6;
Qy 299 QLIHEKLVKPR---IHAVVSGDVNAQTK-----FTQARDFVVKNTIHRVNFV 346
Db 214 ELISFRIIVSGKNVKKLVIGS---LEIENSIDSDFLTQNPVNVLIKIVSDIPISFY 270
Qy 347 NKTIAVAPYLAIDVLVONSQGECEGFRITTEAMAFKLPVLGTAAGTTEIYLDGSTGL 406
Db 271 NN-----MNVFVFPTRREG--FGNVSIQAQALEVPVITTNVTAIDIVVNGEGRF 318
Qy 407 LHPAGKEGVAFLAKNIVRLASHAQRVSMGEGYGRVEMMEHMAERIAVLKQVLK 466

```

```

Db      319 I--VEKGDPRKALAEIKELINDSELAETIGHNGKRVENKFFSQIWEBSMYNTLKE 376
Qy      467 SQ 468
Db      377 SE 378

```

## RESULT 6

```

Y486 MYCLE ID Y486 MYCLE STANDARD; PRT; 428 AA.
AC P54138; O9CB50;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ML2443.
GN ML2443 OR U2168F OR B2168_C2_201.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
   Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,
   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
   Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
   Rutter S., Seeger K., Simon S., Simmonds M., Skellon J., Squares R.,
   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
   RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -1- SIMILARITY: TO M.TUBERCULOSIS RV0486.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way
   modified and this statement is not removed. Usage by and for commercial
   entities requires a license agreement (See http://www.isb-sib.ch/announce/
   or send an email to license@isb-sib.ch).
CC EMBL; U00018; AAA17228.1; ALT_INIT.
DR EMBL; AL583925; CAC31960.1; -.
DR Lepoma; ML2443; -.
DR InterPro: IPR001296; Glycos transf_1.
DR Pfam: PF00534; Glycos transf_1.1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 428 AA; 45291 MW; A14F9F0187B3587C CRC64;

```

```

Query Match 5.0%; Score 119.5; DB 1; Length 428;
Best Local Similarity 21.0%; Pred. No. 0.063;
Matches 90; Conservative 53; Mismatches 174; Indels 111; Gaps 16;
Qy 54 GRDPAALNTVAVAGSPLGMSKVLVLSHELSLSGFLMLMELAFILRVYSQVWVIT 113
Db 13 GIGDAGGMVAVYLO-SALHLARGIEVEIFTRATASADP-----PIVWA 56
Qy 114 NORSQETNDVTVSLDEHMLNHGVYLPARGQAVDI-----ALKADLVILNTAVAGKMD 168
Db 57 -----PGVLAVNVVAGPPEGDKYDLPQCAFAAGVLRABAAHEPGYD 101
Qy 169 PVUKDHVPKVLKILMMIHMGHGFKEVYXKHLPYAGAMDSHTTELYVNSASTDRK 228
Db 102 IVHSH-----YVLSGQVGMARDRA--VPLVHTA-----HTLAAYNAALADGDA 145

```

```

QY 229 TCMPTVY--VHLGNSKEIMEVEADNVRRLREH----- 261
DB 146 AEPRLRSVGEQVVDADDMIVTDTDEKROLISHTHADPAKIDVHNPVDLDMFRGDR 205
QY 262 -TRESLGVRSDDLFAIINSVSRGGDLFLQAFYQALQIHEKTKPRIHVVGSV 320
DB 206 AARAAIGPLDGNVAVFVGRIQPLAIPDIVRA-----AAKLPQRIIVAGGFS 254
QY 321 NAGTFETQLRDFVVKNTIHDRVHV--NKTALAVALAIDVLVONSQGRGECFRIT 377
DB 255 GGSGLASPGGLVRLADDELGITARVTLPPQSTINLATVFOALVAVPS--YSEFGLVA 311
QY 378 IEHMAFKLPLVLTAGAGTTEIYLDGSTGL--HPAG-----KEGAPLAKNIYR 424
DB 312 VERKOCGTPVVAAGVGLPVAVADGVTGLVFGHNVGMADVDQLLISAGPQARISR 371
QY 425 LA-SHAQ 431
DB 372 AAVVHAQ 379

RESULT 7
Y486 MYCTU STANDARD; PRT; 480 AA.
AC 01152;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein Rv0486.
GN Rv0486 OR MT0504 OR MTC2009.12.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sutcliffe J.E., Taylor K., Whitehead S., Barrett J.B.G.,
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umamya L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utermack T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
RN [1]
RP SIMILARITY: TO M.LEPRAE M12443.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
DR EMBL: 277162; CAB00947.1; -
DR EMBL: AE006951; AAK44727.1; -.

```

```

DR TIGR: MT0504; -.
DR Tuberculosis; Rv0486; -.
DR InterPro: IPR001296; Glycos transf_1.
DR Pfam: PF00534; Glycos transf_1; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 480 AA; 50541 MW; 2134755894A9CCF CRC64;

Query Match 5.0%; Score 119; DB 1; Length 480;
Best local similarity 20.8%; Pred. No. 0.081;
Matches 95; Conservative 61; Mismatches 178; Indels 122; Gaps 18;

QY 47 ALRDECDGRG-----DPAALNTAASGSPLGFMRSKVLVLSHLSLS-----G 90
DB 14 AARRRVNGBEATSRSGSPGSRNVSAADP-----RRVALLAVHTSPPLAQGTGADAG 67
QY 91 GPILLMELATLHVNSQVWITNORSGEINDVYTSLSHRLNNGVQVLPANGQEAVIDA 150
DB 68 MNVYMLQSLALHARRGIEVEIFTRATASADPV-----VAVAG----- 106
QY 151 LKADLVILNTRAVAGKW--LDPVLDKDVHPKVLPKILMW-----IHE-----MRGHYF-- 194
DB 107 -----VLRVNVVAGPPEGLD--KYDLPTQLCAFAGVLRABAVHEPGYDIHSHYMLS 158
QY 195 -KVEYVYKLPVYVAGAMIDSHTTAEYMSKTSRDLKIOMQPTVYVHLGNSKELME-----V 248
DB 159 GGQVGLARDRAVPLVHTATLAAVKNALADG--DGEPLPRTVGEQVVDADRLIV 215
QY 249 AEDNVARRVLRH-----IRESLGVRSDDLFAIINSVS 282
DB 216 NNDDEARQVLSLHGADPARIDVHVGVDLDFRPGDRRAARALGLPVDREVVAFAVGRIQ 275
QY 283 RKGQDLFLQAFYQALQIHEKTKPRIHVVGVSDVNAQTKETQLRPFVVKNTIHDR 342
DB 276 PLKADIVLIR-----AAKLPQRIIVAGGFSGLASPDGLVRLADELSNR 324
QY 343 VHFV--NKTALAVALAIDVLVONSQGRGECFRITTBMAFKLPVLTAGAGTTEIYL 400
DB 325 VTFPLPQSHDTLATLPRADIVAVPSY--SEFGLVAVAQAQCPVVAAGLPAVR 382
QY 401 DGSSTGL--HPAGKGVAPLAKNIIVRLAASHAQRV 433
DB 383 DGTITGLVSGHEVQ--WDADIDHLRLCGPRGRV 416

RESULT 8
RFAG ECOLI
ID RFAG ECOLI STANDARD; PRT; 374 AA.
AC P25740;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipopolysaccharide core biosynthesis protein rfaG
DE (Glucosyltransferase I).
GN RFAG OR WAAG OR PCSA OR B3631.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92121133; PubMed=1732225;
RA Parker C.T., Pradel E., Schnaitman C.A.;
RT "Identification and sequences of the lipopolysaccharide core
RT biosynthesis genes rfaG, rfaP, and rfaG of Escherichia coli K-12."
RL J. Bacteriol. 174:930-934 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."

```

RL Nucleic Acids Res. 22:2576-2586(1994).

RN (3)

RP SEQUENCE OF 1-217 FROM N.A.

RX MEDLINE=93077458; PubMed=1447141;

RA Clementz T.;

RT "The gene coding for 3-deoxy-manno-octulosonic acid transferase and the rfaG gene are transcribed from divergently arranged promoters in *Escherichia coli*."

RT *Escherichia coli*;

RL J. Bacteriol. 174:7750-7756(1992).

RN (4)

RP SEQUENCE OF 1-58 FROM N.A.

RX MEDLINE=95080611; PubMed=7988890;

RA Philipinec E., Huismen T.T., Willemsen P.T., Appelmeik B.J., Graef F.K., Oudega B.;

RT "Identification by Tn10 transposon mutagenesis of host factors involved in the biosynthesis of X99 fimbriae of *Escherichia coli*:"

RT effect of LPS core mutations."

RL FEMS Microbiol. Lett. 123:201-206(1994).

CC -1- FUNCTION: INVOLVED IN THE ADDITION OF THE FIRST GLUCOSE RESIDUE TO THE LIPOLYSACCHARIDE CORE.

CC -1- PATHWAY: LIPOLYSACCHARIDE CORE BIOSYNTHESIS.

CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M80599; AAA24082.1; -

DR EMBL; M86305; AAA03743.2; -

DR EMBL; U00039; AAB1608.1; -

DR EMBL; A8000440; AA76655.1; -

DR EMBL; S75736; AAD43826.1; -

DR PIR; B42595; B42595.

DR PIR; B42595; B42595.

DR EcoGene; E01339; rfaG.

DR InterPro; IPR001296; Glycosyltransf\_1.

DR Pfam; PF00534; Glycosyltransf\_1; 1.

KM Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase;

KM Complete proteome.

SO SEQUENCE 374 AA; 42284 MW; 7720AE577CBB80C2 CRC64;

Query Match 4.9%; Score 118; DB 1; Length 374;

Best Local Similarity 24.2%; Pred. No. 0.069;

Matches 50; Conservative 40; Mismatches 91; Indels 26; Gaps 9;

QY 259 REHRESLGVRSDDLFAINSVSRGCGDLFLQAFYQALDLOIHEKLVPRHVVVGS 318

DB 103 REIRCKNGIKKQNLLOVSDGKRGVDRSIEHLASLPESLHNTL-----LFVVGQ 236

QY 319 DVNAQTETQLRDVFNKNTIHDRVFNKTLAVAPYLAIDVLYVNSGSGECFGRIT 378

DB 237 D--KPRFEA---LAETKIGVRSNVHFSGRNDVSELAADLHLHPAY--GEAGIYLL 288

QY 379 EAMAFKLPVGTAAAGTEIVLDGSTG--LLHPAKKEGVNAIAKNVRLA--SHAQRISM 435

DB 289 EAIKAGLPVLTNAGVYAHYADANGCTYIAEPFSGOUL-----NEVLRKALTSPLKAW 344

QY 436 GE--KXGKGVKEMEHMAERIAAVL 460

DB 345 AENRHHYADTDLY--SLPEKADII 368

RESULT 9

VI PC SALTI

ID VI PC SALTI STANDARD; PRT; 578 AA.

AC 004875;

DT 01-NOV-1995 (Rel. 32, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Vi polysaccharide biosynthesis protein vipc/tvie.

GN VIPC OR TVIE OR STY4656.

OS *Salmonella typhi*.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC *Salmonella*.

OK NCBI\_TaxId=601;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=Ty2;

RX MEDLINE=94069051; PubMed=8248629;

RA Maxin H., Virlogeux I., Kolyva S., Popoff M.Y.;

RT "Identification of six open reading frames in the *Salmonella enterica* subsp. *enterica* ser. *typhi* viab locus involved in Vi antigen production."

RT Res. Microbiol. 144:363-371(1993).

RL (2)

RN SEQUENCE FROM N.A.

RP STRAIN=GIPIV 10007;

RC MEDLINE=93322324; PubMed=8331073;

RX Hashimoto Y., Li N., Yokoyama H., Ezaki T.;

RA "Complete nucleotide sequence and molecular characterization of viab region encoding Vi antigen in *Salmonella typhi*."

RT J. Bacteriol. 175:4456-4465(1993).

RL (3)

RN SEQUENCE FROM N.A.

RP STRAIN=CT18;

RC MEDLINE=21534947; PubMed=11677608;

RX Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J., Churcher S., Mungall K.L., Bentley S.D., Holden M.T.G., Sebatina M., Baker S., Basham D., Brooks K., Chillingworth T., Connor P., Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J., Felwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K., Krogh A., Larsen T.S., Leather S., Moule S., O'Goza P., Parry C., Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;

RT "Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar *Typhi* CT18."

RL Nature 413:848-852(2001).

CC -1- PATHWAY: Vi polysaccharide biosynthesis.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X67785; CAA47994.1; -

DR EMBL; D14156; BAA03195.1; -

DR EMBL; A1627283; CAD06776.1; -

DR InterPro; IPR001296; Glycosyltransf\_1.

DR Pfam; PF00534; Glycosyltransf\_1; 1.

KM Complete proteome.

FT CONFLICT 290 S -> G (IN REF. 1).

FT CONFLICT 362 KH -> ND (IN REF. 1).

SO SEQUENCE 578 AA; 65009 MW; 8D420563D86C189 CRC64;

Query Match 4.9%; Score 118; DB 1; Length 578;

Best Local Similarity 22.2%; Pred. No. 0.13;

Matches 69; Conservative 58; Mismatches 132; Indels 52; Gaps 14;

QY 155 LVLTNNAVAGKMDPVAKDHVPRV-----LPKLTWIEHMGHFKVEY---VKHLPRV 205

DB 294 WMLAALING-----VPRQLGIRGLPRV-----VRKLFKEVEPRLYQALAVV 337

QY 206 AGA--MDSHTTAEVWNSRSDRLKIQPOTVVVHLGNSKEIMEVAEDNVARVLEHIR 263

DB 338 PGVDFEWSNNHCSTRHY---ADWLKLEAKHPQVVYNGVLPSTEPSSE--VPHKIKWQGFQ 392

QY 264 ESIQVRSDDLFAINSVSRGCGDLFLQAFYQALDLOIHEKLVPRHVVVGS--SDVNA 322

DB 393 KI---QDAD---TTIGGVFRFVGDKNPFIWIDFARYLQHH---PATRFVLYGDGDLRA 442





CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

[illegible]

Qy	8	AAVAAVAGRGVNHRRQLLOLLLLVAAVAAASVSTGFL--	RGALRDCGCGRDPALINTA-	64	
Db	1107	AOBMAAGSGEGFLP----	SLEBNDALARRNAKKEVDQREDDYATIVAA	1152	
Qy	65	-----VASGSPFLG-----	-----FMRSKL	78	
Db	1153	SEALLAAAGATGELGALDEBWLPHLELGMHKKLLGMEARREALVONAIYOLFRLDRLOAL		1212	
Qy	79	VLLVSHEDLSLGGGL-----	LLMELATFLRHVGSQVWITNORSQPTNDVITYSLE	128	
Db	1213	VLVRRGWAALSSGALLEGTVSYEEALKQNRHDL-----	TTMELSSQKQVAVQQA	1262	
Qy	129	HRRLAHGVOVLPARGENDVIALQADVLITANVAGKVL-----	DPLKDDHVPKYLEKILM	184	
Db	1263	EGALLRG--NIVGROAGEAVATRLERKNE--	NOLRRQOMKQKLDQLELQHLRDCHELDG	1319	
Qy	185	WHEKMGHYFVFEVYGHLEFPVAGAMIDSHTTAEVYNSRSTDRKLTIOMQPTVYVHLLGNSKE		244	
Db	1320	WIHE-----	QMLMARGDREDDHKKHKKLH-----	QAFMELAQNKE	1358
Qy	245	LMVAVEDVARRVLR-----	HLRESLG-VNSBELLFALINSVGRGQDLF-----	290	
Db	1359	WLEKLE-REGQOLMOEKREPLAASVRKKLSTGRQ--	CAALSTTQAKARQLEPASKADQ	1414	
Qy	291	LDQAFYQALQLOLHEKLVKPRHIAHVVGSD--	VNAQTK-----	FETQLRDPVYKNTIHDR	342
Db	1415	LVQSPAEIDKLLHMR-----	SOLQDVPDGDGLATVNSOLKLGSMESQVEW-----	YRE	1465
Qy	343	VHFVQKTLTAVPYLAIDVLY-----	QNSQSG	368	
Db	1466	VGELOAQTRALPLEPSKELVGERQNAVQ		1494	
RESULT	12				
ID	UGST	MAIZE	STANDARD;	PRT; 605 AA.	
AC	P04713;				
DT	13-AUG-1987	(Rel. 05, Created)			
DT	13-AUG-1987	(Rel. 05, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	Granule-bound glycogen [starch] synthase, chloroplast precursor				
GN	WAXY.				
OS	Zea mays (Maize).				
OS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;				
OC	Panicoidae; Andropogoneae; Zea.				
OX	NCBI_TaxID=4577;				
RN	[1]				
RA	SEQUENCE FROM N.A.				
RP	Kloessen R.B., Gierl A., Schwarz-Sommer Z., Sedler H.;				
RT	Molecular analysis of the waxy locus of Zea mays."				
RL	Mol. Gen. Genet. 203:237-244(1986).				
CC	-1- FUNCTION: REQUIRED FOR THE SYNTHESIS OF ANYULOSE IN ENDOSPERM.				
CC	-1- CATALYTIC ACTIVITY: UDP-glucose + {(1,4)-alpha-D-glucosyl} (N) =				
CC	UDP + {(1,4)-alpha-D-glucosyl} (N+1).				
CC	-1- PATHWAY: Starch biosynthesis.				
CC	-1- SUBCELLULAR LOCATION: CHLOROPLAST; GRANULE-BOUND.				
CC	-1- SIMILARITY: BELONGS TO THE BACTERIAL/PLANT GLYCOGEN SYNTHASE				
CC	FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.ebi.ac.uk/announcements">http://www.ebi.ac.uk/announcements</a>				
CC	or send an email to <a href="mailto:license@ebi.ac.uk">license@ebi.ac.uk</a> ).				
CC	-----				
CC	EMBL; X03935; CAA27574.1; -				
CC	EMBL; M24258; AA333520.1; -				
CC	PIR; S07314; S07314.				

DR MaizeDB; 15806; -, Glycos transf. 1.  
 DR InterPro; IPR001296; Glycos transf. 1.  
 DR Pfam; PF00534; Glycos transf. 1; 1.  
 KW Glycogen biosynthesis; Transferase; Glycosyltransferase;  
 FT TRANSIT 1 72 CHLOROPLAST; Starch biosynthesis.  
 FT TRANSIT 73 605 GRANULE-BOUND GLYCOGEN (STARCH) SYNTHASE.  
 FT BINDING 91 91 UDP-GLUCOSE (BY SIMILARITY).  
 SQ SEQUENCE 605 AA; 65966 MW; 137F15207DEFC189 CRC64;

Query Match 4.5%; Score 108.5; DB 1; Length 605;  
 Best Local Similarity 22.8%; Pred. No. 0.75; Indels 41; Gaps 7;  
 Matches 44; Conservative 35; Mismatches 73;

QY 254 ARRVLRHRSLSG--VRSDDLPAIINSVSRGGDLPQAFYALQIQHEKLVPR 311  
 DB 374 AKAKKEALQAEVGLPYDRNIPLVAFIGRLERQKGPVMAAAIPQIMEMVED-----V 426  
 QY 312 HAVVVGSDVNAQTFFETQLADPVVKNTHDRHFVNKTLAVAPYLAAL-----DVLV 363  
 DB 427 QIVLIGT---GKKFFEMLSAEK-----PRGVRAVVKFMAALAHIMAGDVLA 475  
 QY 364 QNSQGRGRCGRITTEMAFKLPVLGTAGTTEIVLDSTG-----LHPACK 412  
 DB 476 VTS--RFEPGCLIQOGMRGTFCACASTGLVDTIIEGKTGFHMRSLVDVCNVEPADV 533  
 QY 413 EGVAFLAKNIVRL 425  
 DB 534 KKVATTIQRAIKV 546

RESULT 13  
 LPSE RHIME STANDARD; PRT; 340 AA.  
 ID LPSE RHIME STANDARD; PRT; 340 AA.

AC 09R9N1;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Lipopolysaccharide core biosynthesis glycosyl transferase lpse  
 DE (EC 2.-.-.-).  
 GN LPSE OR R01571 OR SMC01220.  
 OS Rhizobium meliloti (Sinorhizobium meliloti).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Sinorhizobium.  
 OC NCBI\_TaxId=382;

RP SEQUENCE FROM N.A.  
 RC STRAIN=RCR2011 / S047;  
 RA Lagares A., Hozbor D.F., Niehaus K., Pich Otero A.U.L., Lorenzen J.,  
 RA Arnold W., Puehler A.;  
 RT "Genetic characterization of a Sinorhizobium meliloti chromosomal  
 RT region involved in lipopolysaccharide biosynthesis.";  
 RL J Bacteriol. 183:1248-1258(2001).

RN [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=1021;  
 RX MEDLINE=2136507; PubMed=11481430;  
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,  
 RA Boletard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,  
 RA Godrie T., Goffeau A., Kahn D., Kles A., Leliane V., Masny D.,  
 RA Pohl T., Portefelle D., Puehler A., Purnelle S., Rambert U.,  
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Gilbert F.;  
 RT "Analysis of the chromosome sequence of the legume symbiont  
 RT Sinorhizobium meliloti strain 1021.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).  
 CC -I- PATHWAY: Lipopolysaccharide core biosynthesis.  
 CC -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 1.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; AF193023; AAF06009.1; -;  
 DR EMBL; AL591787; CAC46150.1; -;  
 DR InterPro; IPR001296; Glycos transf. 1.  
 DR Pfam; PF00534; Glycos transf. 1; 1.  
 KW Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase;  
 KW Complete proteome.  
 SQ SEQUENCE 340 AA; 38036 MW; D8363F95FD9E169F CRC64;

Query Match 4.5%; Score 107; DB 1; Length 340;  
 Best Local Similarity 19.8%; Pred. No. 0.44; Indels 114; Gaps 17;  
 Matches 76; Conservative 53; Mismatches 140;

QY 130 RMLNHGVQLPARGQAVNDIA-----LKADLVLTNTAVAGK 166  
 DB 16 RFPVHLVNALAEVGTQATIRPGWRDIEGAKIRESHFRMLSIDRILL----- 67  
 QY 167 LDPVTKOHV-----PVVLPKILMW---IHMRGHY---FKVEYVGHLPFAAGMIDSHHT 215  
 DB 68 --PLKVKHMAARRKRPVVL---MAMAPRASLMEPNYKGAFKISRGDYF-----TR 112  
 QY 216 AAYWNSRTSRLKIQNP--QTYVHLGNSKELMEVAEDNVARVLRHIRESLGVRSDDL 273  
 DB 113 LSYF--RNTDICIYCNTPGIAERVSDLGKREIRIVISNFTGTGRVV----- 155  
 QY 274 LPAIINSVSRGK-----GODLPQAFYALQIQHEKLVPRHIAVVGSDV 320  
 DB 156 -----NVDKAKLDTADAPVVMGKRFVERKGFHTLEAV---ALPQVYIMLIGGE 205  
 QY 321 NAQT--KEFTQLRDPVVKNTIHDRVFNKTLAVAPYLAALDVLVQNSQGRGRCGRIT 378  
 DB 206 EEDNLHKLAILDL-----GVSGVRPAGQDDTRPFLAADVVMSS--HEPLGNVIL 256  
 QY 379 EAMAKLPVLGTAGTTEIVLDSTGLHPAG--KEGVAPLAKNIIVLASHAQRVSMKE 437  
 DB 257 ESMAGTIPVSTSESPQWFMGDGNGMLAVDTGDAG--PARALEQIVADNLSRLAE 313  
 QY 438 KQYGRVXEMFEMHMAERLAVAL 460  
 DB 314 RGHETLVQFSRBAITDAVLIQL 336

RESULT 14  
 USST HORVU STANDARD; PRT; 603 AA.  
 ID USST HORVU STANDARD; PRT; 603 AA.

AC P09842;  
 DT 01-MAR-1989 (Rel. 10, Created)  
 DT 01-MAR-1989 (Rel. 10, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Granule-bound glycogen [starch] synthase, chloroplast precursor  
 DE (EC 2.4.1.11).  
 GN WAXY.

OS Hordeum vulgare (Barley).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;  
 OC Triticeae; Hordeum.  
 OC NCBI\_TaxId=4513;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv Vogelzanger Gold; TISSUE=Leaf;  
 RX MEDLINE=8830345; PubMed=2970062;  
 RA Rhode W., Becker D., Salamin F.;  
 RT "Structural analysis of the waxy locus from Hordeum vulgare.";  
 RL Nucleic Acids Res. 16:7185-7186(1988).  
 RN [2]

RP SEQUENCE OF 76-89.  
 RC STRAIN=cv H354-295-2-5; TISSUE=Starchy endosperm;  
 RX MEDLINE=94170739; PubMed=8125056;  
 RA Flengerd R.;  
 RT "Separation of acidic barley endosperm proteins by two-dimensional



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: June 30, 2003, 16:18:51 ; Search time 36 Seconds  
(without alignments)  
2707.230 Million cell updates/sec

Title: US-09-938-294-45

Perfect score: 2397  
Sequence: 1 MAKTPSFVAVAAGRGPHV.....ERIAAVLKDVRKSGHSHS 473

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

1: SP\_archaea:\*  
2: SP\_bacteria:\*  
3: SP\_fungi:\*  
4: SP\_human:\*  
5: SP\_invertebrate:\*  
6: SP\_mammal:\*  
7: SP\_mhc:\*  
8: SP\_organelle:\*  
9: SP\_phage:\*  
10: SP\_plant:\*  
11: SP\_prodent:\*  
12: SP\_virus:\*  
13: SP\_vertebrate:\*  
14: SP\_unclassified:\*  
15: SP\_virus:\*  
16: SP\_bacteriap:\*  
17: SP\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1478	61.7	402	10 Q9FWT0	Q9FWT0 arbidopsin
2	1471	61.4	458	10 Q9FXG9	Q9FXG9 arbidopsin
3	671	28.0	188	10 Q9LI48	Q9LI48 oryza sativ
4	487	20.3	697	10 Q9LSB5	Q9LSB5 arbidopsin
5	482.5	20.1	670	10 Q9SSP6	Q9SSP6 arbidopsin
6	207	8.6	382	16 Q9UAI5	Q9UAI5 agrobacteri
7	197	8.2	416	17 Q9H762	Q9H762 pyrococcus
8	187.5	7.8	388	16 Q9WZ90	Q9WZ90 thermotoga
9	182.5	7.6	424	5 Q9SSA2	Q9SSA2 drosophila
10	181.5	7.6	381	17 Q9S512	Q9S512 pyrococcus
11	180	7.5	412	17 Q9H000	Q9H000 pyrococcus
12	179.5	7.5	424	5 Q9VZU8	Q9VZU8 drosophila
13	172.5	7.2	370	16 Q9PEZ9	Q9PEZ9 xyella fas
14	169.5	7.1	381	2 Q9EVX4	Q9EVX4 streptococc
15	166	6.9	382	16 Q9YCW3	Q9YCW3 anabaena sp
16	165	6.9	471	3 Q9X0H8	Q9X0H8 neurospora

17	163.5	6.8	360	17 Q8Z246	Q8Z246 pyrobaculum
18	163	6.8	404	16 Q8RCY0	Q8RCY0 thermoaer
19	163	6.8	422	16 Q98J63	Q98J63 rhizobium 1
20	162.5	6.8	385	2 Q9X4V1	Q9X4V1 streptococc
21	161	6.7	380	16 Q8RBZ6	Q8RBZ6 thermoaer
22	161	6.7	405	10 Q9ZV98	Q9ZV98 arbidopsin
23	157.5	6.6	506	3 Q13604	Q13604 schizosacch
24	157	6.5	511	3 Q96WV6	Q96WV6 schizosacch
25	156	6.5	416	16 Q9ZVR7	Q9ZVR7 rhizobium m
26	155	6.5	290	2 Q8WV73	Q8WV73 pasteurella
27	154.5	6.4	351	17 Q58577	Q58577 methanococc
28	154	6.4	358	17 Q8TZU8	Q8TZU8 pyrococcus
29	154	6.4	422	2 Q937E1	Q937E1 nostoc punc
30	151.5	6.3	411	17 Q26550	Q26550 methanobact
31	151.5	6.3	414	16 Q9ZQ88	Q9ZQ88 rhizobium m
32	151	6.3	375	16 Q9KC90	Q9KC90 bacillus ha
33	150	6.3	406	16 Q8RBZ4	Q8RBZ4 thermoaer
34	149.5	6.2	393	16 Q9WZ95	Q9WZ95 thermotoga
35	149	6.2	381	16 Q8Z2F9	Q8Z2F9 salmoneila
36	148.5	6.2	378	16 Q9HUG1	Q9HUG1 pseudomonas
37	148.5	6.2	383	2 Q9RHD1	Q9RHD1 pseudomonas
38	148.5	6.2	436	16 Q9XYG6	Q9XYG6 streptococ
39	147.5	6.2	351	16 Q8UDE4	Q8UDE4 agrobacteri
40	147.5	6.2	380	2 Q88201	Q88201 escherichia
41	147	6.1	333	17 Q9HSV4	Q9HSV4 halobacteri
42	145	6.0	395	16 Q9K6L7	Q9K6L7 bacillus ha
43	144	6.0	409	16 Q55598	Q55598 synechocyst
44	144	6.0	425	2 Q939V0	Q939V0 anabaena sp
45	144	6.0	429	16 Q8YRS3	Q8YRS3 anabaena sp

## ALIGNMENTS

RESULT 1	Q9FWT0	PRELIMINARY;	PRT;	402 AA.
ID	Q9FWT0			
AC	Q9FWT0			
DT	01-MAR-2001 (TREMBL)	16, Created		
DT	01-MAR-2001 (TREMBL)	16, Last sequence update		
DT	01-JUN-2001 (TREMBL)	17, Last annotation update		
DE	F1B16.5 protein.			
GN	F1B16.5			
OS	Arbidopsin thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid1; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
RN	SEQUENCE FROM N.A.			
RP	Federici N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,			
RA	Altati H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,			
RA	Buehler E., Chao Q., Chin C., Chou J., Choi E., Gonzalez A.,			
RA	Hwang B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,			
RA	Ianz C., Liu A., Liu S., Mukharly N., Pham P., Sakano H., Shinn P.,			
RA	Torimaru M., Vayenberg M., Yu G., Ecker J., Theologis A., Davis R.W.,			
RU	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AC023754; AAC13070.1; -			
DR	InterPro; IPR001296; Glycos transf_1.			
DR	Pfam; PF00534; Glycos transf_1; 1.			
DR	SEQUENCE 402 AA; 44901 MW; 5820117CD4E13DD5 CRC64;			
SO	SEQUENCE			
Query Match	61.7%;	Score 1478;	DB 10;	Length 402;
Best Local Similarity	74.5%;	Pred. No. 1.6e-107;		
Matches 283;	Conservative 53;	Mismatches 44;	Indels 0;	Gaps 0;
QY	89 SGGPLLMELAFILRHVSGQVWITNROSOETNDVYSLERHMLHGVOLPARGQEAVID 148			
DB	18 SRGPLLMEIAFLRLGVAGADVWITNCKPLEDDVYSLERHMLHGVOLPARGQEAVID 77			
QY	149 IALKADVITINAVAGKWLDPVTKDHPVPLPKITLWTHMRGHVYEVVGHLPFVAGA 208			
DB	78 TSLKADVITINAVAGKWLDPVTKDHPVPLPKITLWTHMRGHVYEVVGHLPFVAGA 137			

QY 209 MIDSHTAAYNSRTSDRLKIQMPOYVYVHLGNSKELMEVAEDNVARVLEBHIRESLGV 268  
 DB 138 MIDSHATAGYWKNGKQARLGKIPKTYVHLGNSKELMEVAEDSVAKRYLREHRESLGV 197  
 QY 269 RSEDLPAIINSVSGKQODLFLQAFYQALQIOLHEKIKVPRIHAVVSGDVNAQTFET 328  
 DB 198 RNEBLLFGIINSVSGKQODLFLRAFHESLERIKKQVPTMAHAVVSGDMSKQTFET 257  
 QY 329 QLRDFVVKNTIDRVHFVNKTLAVAPYLAIDVLVONSQSGRCFCRITTEMAFKLPVL 388  
 DB 258 ELNFPREKKLENFHFVNKTLVAPYLAIDVLVONSQSGRCFCRITTEMAFKLPVL 317  
 QY 389 GTAAGTTEIVLDSTGLHPAGKEGVAFLANNVRLASHAEQVSMGEKGYKEMEM 448  
 DB 318 GTAGAGTMEIVNGTGLHSAGEGVPLANNIKVATQVELRLMGKNGYERKEMFL 377  
 QY 449 EHNMAERIAVLDVLRKSQ 468  
 DB 378 EHNMSHRISVLRKVLQNAK 397

## RESULT 2

Q9FXG9 PRELIMINARY; PRT; 458 AA.  
 ID Q9FXG9  
 AC Q9FXG9  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE F6P9.24 protein.  
 GN F6P9.24  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 NX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Federle H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,  
 RA Buehler E., Chao Q., Chou J., Choi E., Gonzalez A.,  
 RA Hwang B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,  
 RA Lenz C., Liu A., Liu S., Mukharezy N., Pham P., Sakano H., Shinn P.,  
 RA Toriumi M., Vayberg M., Yu G., Ecker J., Theologis A., Davis R.W.,  
 RA Submitted (SRP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AAC007797; AAC12556.1  
 DR InterPro; IPR001296; Glycos\_transf\_1.  
 DR Pfam; PF00534; Glycos\_transf\_1; 1.  
 SQ SEQUENCE 458 AA; 51367 MW; 609A1B9DBB818740 CRC64;

## Query Match

Best Local Similarity 62.8%; Score 1471; DB 10; Length 458;  
 Matches 304; Conservative 65; Mismatches 73; Indels 42; Gaps 6;

QY 1 MAKTPSPAAYAAVAGRGVPHNRTOQLLLVAVASASTAGFLLRQALRDPDGRG---- 56  
 DB 1 MAKTPSPMAATLQKKRMP-----LMLLV---LSTVGMILVSTDSGVSGRCS 50  
 QY 57 -----DPAALNTAVASGS--PLGFWRSLVLYVSHSLISGGPILLMELAFLLRHVGSQV 110  
 DB 51 REKEDNSDIKIOSVGSGLNPLFEPWKSIVLVSHSLISGGPILLMELAFLLRGVSEVY 110  
 QY 111 WITNQSQSTNDVYSLSHRMLNHGVOVPRAGQANDYALKVDYILTANTAVAGMDPV 170  
 DB 111 WITNQSQSTNDVYSLSHRMLNHGVOVPRAGQANDYALKVDYILTANTAVAGMDPV 170  
 QY 171 LKQHVPRVLPKILMWHIEMRGHYEKVEYKHLPPVAGAMIDSHTTAEYNSRTSDRLKIQ 230  
 DB 171 LKQHVPRVLPKILMWHIEMRGHYEKVEYKHLPPVAGAMIDSHTTAEYNSRTSDRLKIQ 230  
 QY 231 MPQYVYVHNGSKELMEVAEDNVARVLEBHIRESLGVRSDDLPAIINSVSGKQODL 230  
 DB 231 MPQYVYVHNGSKELMEVAEDNVARVLEBHIRESLGVRSDDLPAIINSVSGKQODL 230

QY 291 LQAFYQALQIOLH-EKIKVPRIHAVVSGDVNAQTFETQLRDFVVKNTIDRVHFVNKT 349  
 DB 270 LRFHESLKVKEKLEVPPTMAHAVVSGDMSKQTFETELRNFPQEMKQKQIVHFVNKT 329  
 QY 350 LAVAPYLAIDVLVONSQSGRCFCRITTEMAFKLPVLGTAAGTTEIVLDSTGLHP 409  
 DB 330 MKVAPYLAIDVLVONSQSGRCFCRITTEMAFKLPVLGTAAGTTEIVYVNTTGLLN 389  
 QY 410 AKREGVAPLANIVRLASHAEQVSMGEKGYKEMEMFHNMAERIAAVLDVLRKSQ 469  
 DB 390 TKGDDVPLANNIKVATQVELRLMGKNGYERKEMFLDEHNMSHRISVLRKVLQNAK 449  
 QY 470 HSRK 473  
 DB 450 HSRK 453

## RESULT 3

Q9LI48 PRELIMINARY; PRT; 188 AA.  
 ID Q9LI48  
 AC Q9LI48  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE EST C74729 (B50675) corresponds to a region of the predicted gene.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Eriatoidae; Oryzaceae; Oryza.  
 NX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Macsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 RT clone: P0469E09."  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF001366; BAA92419.1; -  
 DR InterPro; IPR001296; Glycos\_transf\_1.  
 DR Pfam; PF00534; Glycos\_transf\_1; 1.  
 SQ SEQUENCE 188 AA; 20622 MW; C7C4E168752107DA CRC64;

## Query Match

Best Local Similarity 91.6%; Score 671; DB 10; Length 188;  
 Matches 131; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

QY 311 IHAUVSGDVNAQTFETQLRDFVVKNTIDRVHFVNKTLAVAPYLAIDVLVONSQSGRG 370  
 DB 1 MHAUVSGDVNAQTFETQLRDFVVKNTIDRVHFVNKTLAVAPYLAIDVLVONSQSGRG 60  
 QY 371 ECGRRTTEMAFKLPVLGTAAGTTEIVLDGSGTGLHPAGKEGVAFLANNVRLASHAE 430  
 DB 61 ECGRRTTEMAFKLPVLGTAAGTTEIVLDGSGTGLHPAGKEGVAFLANNVRLASHAE 120  
 QY 431 QRVSMGEKGYKEMEMFHNMA 453  
 DB 121 DRVSMGEKGYKEMEMFHNMA 143

## RESULT 4

Q9LSB5 PRELIMINARY; PRT; 697 AA.  
 ID Q9LSB5  
 AC Q9LSB5  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Glycoyl transferases-like protein (AT3g15940/MYC8\_7).  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 NX NCBI\_TaxID=3702;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RL Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;  
Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=20277480; PubMed=10819329;  
RA Nakamura Y.,  
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence  
features of the regions of 4,504,864 bp covered by sixty P1 and TAC  
RT clones";  
RN [3]  
RL DNA Res. 7:131-135(2000).  
RP SEQUENCE FROM N.A.  
RA Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,  
Bower L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,  
Hayshtizaki Y., Ishida J., Jones T., Kamuya A., Karlin-Neumann G.,  
Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,  
Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,  
Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,  
Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,  
Ecker J.R.;  
RT "Arabidopsis cDNA clones";  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB026653; BAB02880.1; -;  
DR InterPro; IPR001199; Cyt B5.  
DR InterPro; IPR001296; Glycos transf\_1.  
DR Pfam; PF00534; Glycos transf\_1; 1.  
DR PROSITE; PS00191; CYTOCHROME\_B5\_1; UNKNOWN\_1.  
KW Transfease.  
SQ SEQUENCE 697 AA; 78706 MW; 1732545669F315 CRC64;  
Query Match 20.3%; Score 487; DB 10; Length 697;  
Best Local Similarity 29.1%; Pred. No. 1.7e-29;  
Matches 147; Conservative 77; Mismatches 159; Indels 122; Gaps 13;  
QY 52 CDORGPALMTAVASGSPGPNR---SKVLVLSHLSLGGPPLLMELFLRHGSG 108  
DB 223 CDKSPD-----FRLVMSRRFVLPHLSMTGAPLSMELSELSCGAT 267  
QY 109 VVWITNQSQETNDVTVSLSHRMNHGVQVLPARGOAVDIALKDLVLTAVAGKMD 168  
DB 268 VVAVVLSSRG-----GLGDELTRRIKIVVEDGSELPFTAKMDVLVAGSVCASWID 320  
QY 169 PVLKDVPRVPLKILMITHMRGHRYK-----VEVYKLPFVAGAMDSHTAAYNMS-R-222  
DB 321 QYV-DHHPAGGSQILAMWMMNRREYDRAKPVLRVXLILF-----SEVQSKMLTWC 373  
QY 223 TSDRLKIOMPQTVVYVHLGNSKEIMEVAEDNVA-----RVYLRHETRESLAV 268  
DB 374 EEDHVYKLR-SQPVIVPLSVNDELAFVAGSSSLMTPTLTQETWKEKROKLESVRTEFGL 432  
QY 269 RSEDLLFALINSVSKGKQDLFLQATYQALQ-----IQHKKK 307  
DB 433 TDDKMLVMSLSSINSNGQLLLESVALLEREQTQOVAKRNQSIKIKNIGIRKSKIS 492  
QY 308 VPRIHAV-----VVGSDVA 321  
DB 493 LSAHRHLRGSSRKMKITSPAVDNHPSVLSATGRRRLKLLSGNTQKODKLKLLSVGSKSN 552  
QY 322 AQCKFEQQLRDFVYKVN-TIDRHYFNPKTLAAAPLALIVLVYNSQGRGECFRITTEA 380  
DB 553 -KVAAYVKEMLSPSSNNGLSNVLMTPATRRVSLVSAADVYTNNSQGGELFRRVITEA 611  
QY 381 MAFKLPLVLTAGATTEIVLDGSTGLHDPAGKEGVAPLAKNIVRLASHAEORVAMGKGY 440  
DB 612 MAYGLPVLGTADAGCKEIVHENYVGLHHPVGAQNKVLAQNLLFLRNPSTRRLQSGQGR 671  
QY 441 GRVKEFMENHMAEKLAAVLDKDLR 465

DB 672 EIVEKMYMKQMYKRPVDVLYKMR 696

## RESULT 5

Q9SSP6 PRELIMINARY; PRT; 670 AA.

AC Q9SSP6; 01-MAY-2000 (TREMBL) 13, Created)

DT 01-MAY-2000 (TREMBL) 13, Last sequence update)

DT 01-JUN-2001 (TREMBL) 17, Last annotation update)

DE F6D8.36 protein (Glycosyl transferase, putative).

GN F6D8.36 OR F19K6.13.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eucotyledons; Brassicales; Brassicaceae; Arabidopsids.

NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CV, COLUMBIA;

RA Vysotskaia V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,

Lee J.M., Li J., Gonzalez A., Liu A., Liu K., Mukharasy N., Sakano H.,

Vaysberg M., Chin C., Choi E., Chou J., Altafi H., Araujo R.,

Brooks S., Buehler E., Chao O., Conn L., Conway A.B., Dunn P.,

Hansen N., Hwang B., Huizar L., Khan S., Kim C., Palm C., Rowley D.,

Shinn P., Walker M., Davis R.W., Ecker J.R., Federspiel N.A.,

Theologis A.;

RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CV, COLUMBIA;

RA MEDLINE=21016719; PubMed=1130712;

RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,

White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,

Buehler E., Chan A., Chao O., Chen H., Cheuk R.F., Chin C.W.,

Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,

Dunn P., Egu P., Feldblyum T.V., Feng J.-D., Feng B., Fujii C.Y.,

Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,

Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khavkin E.,

Kim C.J., Koo H.L., Kremetskaia I., Kurtz D.B., Kwan A., Lam B.,

Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.-P.,

Lin X., Liu S.X., Liu Z.A., Luros J.S., Malt R., Marzilli A.,

Militschev J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,

Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,

Sakano H., Salzer S.L., Schwartz J.R., Shinn P., Southwick A.M.,

Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,

Utreback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,

Wu D., Yu G., Fraser C.W., Venter J.C., Davis R.W.;

RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis

thaliana";

RL Nature 408:816-820(2000).

DR EMBL; AC008016; AAG55621.1; -;

DR EMBL; AC037424; AAG51540.1; -;

DR InterPro; IPR001199; Cyt B5.

DR InterPro; IPR001296; Glycos transf\_1.

DR Pfam; PF00534; Glycos transf\_1; 1.

DR PROSITE; PS00191; CYTOCHROME\_B5\_1; UNKNOWN\_1.

KW Transfease.

SQ SEQUENCE 670 AA; 75636 MW; F59FE275C33DD0AF CRC64;

Query Match 20.1%; Score 482.5; DB 10; Length 670;

Best Local Similarity 29.2%; Pred. No. 3.6e-29;

Matches 148; Conservative 77; Mismatches 173; Indels 109; Gaps 13;

QY 38 STAFGLR--GALRD-----PCDGRDPALMTAVASGSPGPNR---SKVLV 80  
DB 193 ATYGKLLDPFSSLEDVYKLEWSPHRRSGTCDRKSD-----FRLVMSRRFV 237  
QY 81 IVSHSLSSGGPPLMELFLRHVGSQVWITNQSQETNDVTVSLSHRMNHGVQVLP 140  
DB 238 LFLHLSMTGAPLSMELSELSCGATVSAADVYTNNSQGGELFRRVITEA 290

QY 141 ARGEADVADLAKADVIINTAVAGKMDPVLKDHVETKPKILMWHKMGHYFK----- 195  
 DB 291 DKGEISFKTAMKADLLIAGSAVCTSWIDQVMNH- FAGSGQIAWIMENREYFDRAPV 349  
 QY 196 VEVYVHLPVAGAMIDSHHTAEVYNS-RTSDRLKIQPOTVYVHLGNSKELMEV- 249  
 DB 350 LDRVYMLIFL-----SESQSRQHLTWCEEHKLR-SQPIVYPLSNDLEAFVGISS 402  
 QY 250 -----EDNVARVLEHRIHRSLSGVRSEDLLFAIINSVRKGDPLQAFYQALQ- 299  
 DB 403 INTPLSPERQVRKROILRESVTELEGITSDMLVMSLSINPKGQLLLESIALLASE 462  
 QY 300 -----LIGHEKLVPRINAV-----VVGS 318  
 DB 463 RQGESQRNHGIIIRKKNVSLSKRLGSSRQKMSVSLTDLNGLRKKEKLVLGSGVS 522  
 QY 319 DVNAQTRETOLEDFVVKNTIHDVHFNKTLAVAPYLAIDVLVNSQGRGECFGRIT 378  
 DB 523 KSNKGVYKEMLSFLSNSGNSKSVMTPTATTVASLVSADVYVTVNSQGVETGRTYI 582  
 QY 379 EAMAFKLPVLGTAAAGTTEIVLDGSTGLHPACKEGVAPLAKNIVRLASHABQVSMEEK 438  
 DB 583 EAMAYGLAVGTAGATKEMVQHNMTGLLSMGRSANKELAHNLVILRNPDRLRLOSE 642  
 QY 439 GYGVKEMEMEHMAERIALAVLKDLR 465  
 DB 643 GRKMYEKVMKQHMVKEKFPVDVLVKMR 669

## RESULT 6

QY 08UAI5 PRELIMINARY; PRT; 382 AA.

AC Q8UAI5; TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Glycosyltransferase.  
 GN ATU3560 OR AGR L.2541.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Rhizobiaceae; Rhizobium.  
 NCBI\_TaxID=176299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608550; PubMed=11743193;  
 RA Wood D.W., Setubal J.C., Kaul R., Monke D.E., Kitaajima J.P.,  
 RA Okura Y., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,  
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,  
 RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,  
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan V., Perry M.,  
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Nester E.W.,  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 RT C58."  
 RL Science 294:2317-2323 (2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608551; PubMed=11743194;  
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Mullin L.,  
 RA Querol B., Goldman B.S., Cao Y., Akenazi M., Halling C., Mullin L.,  
 RA Hounel K., Gordon J., Vaudin M., Jarchock O., Epp A., Liu F.,  
 RA Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markfeld B.,  
 RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,  
 RA Cielo C., Slater S.,  
 RT "Genome sequence of the plant pathogen and biotechnology agent  
 RT Agrobacterium tumefaciens C58."  
 RL Science 294:2323-2328 (2001).  
 DR EMBL; AE009285; AA144372.1;  
 DR EMBL; AE008327; AA89838.1;  
 KW Transferase; Complete proteome.

SQ SEQUENCE 382 AA; 40400 MW; 802885D24D0F042A CRC64;  
 Query Match 8.6%; Score 207; DB 16; Length 382;  
 Best Local Similarity 27.5%; Pred. No. 5.9e-08;  
 Matches 68; Conservative 46; Mismatches 109; Indels 24; Gaps 7;  
 QY 220 NSTRDPLKI-----QMPOTVYVHLGNSKELMEVADNVARRVIREHRSIGVRSDDL 275  
 DB 154 NSEETGRAPTEENGEGADVRIVYNGFDPAKKALHDAGMAR-----LRRELGLPQPLV- 207  
 QY 276 AIINSVRKGDPLQAFYQALQLIGHEKLVPRINAVVAVGSDVNAQTEFOLRPFV 335  
 DB 208 GLFERLSKKGQHVFDLAL-ANMEGV-----AVVGAALFQGEAEVATIRQAS 256  
 QY 336 KNTHDRVHFVNKTLAVAPYLAIDVLVNSQGRGECFGRITTEAMAFKLPVLGTAA 395  
 DB 257 RLGLDRVRFVLFGRSDVPELMASMDVVAHTSI-VAEPFGVVEAMMCGRPVAVATRG 315  
 QY 396 TEIVLDGSTGLHPACKEGVAPLAKNIVRLASHAEGQVSMGEKGYGVKEMEMHMAER 455  
 DB 316 TEIRROGEFGLVPPG--DASALAAALGTLISDPAALQRLGSGREDVSDFSIQETCRS 373  
 QY 456 IAVLKD 462  
 DB 374 VSALLTE 380

## RESULT 7

QY 058762 PRELIMINARY; PRT; 416 AA.

AC 058762; TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)  
 DE Hypothetical protein PH1035.  
 GN PH1035.  
 OS Pyrococcus horikoshii.  
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;  
 OC Pyrococcus.  
 NCBI\_TaxID=53953;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OT3;  
 RX MEDLINE=98344137; PubMed=9679194;  
 RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohtoku Y.,  
 RA Funahashi T., Tanaka T., Kudo Y., Yamazaki Y., Kushiida N., Oguchi A.,  
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
 RA Masuchi Y., Shizuya H., Kikuchi H.,  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."  
 RL DNA Res. 5:55-76 (1998)  
 DR EMBL; AF000004; BAA30133.1;  
 DR InterPro; IPR001296; Glycosyltransferase\_1.  
 DR Pfam; PF00534; Glycosyltransferase\_1;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 416 AA; 48196 MW; 89BDBEE51CE7B0 CRC64;

Query Match 8.2%; Score 197; DB 17; Length 416;

Best Local Similarity 25.5%; Pred. No. 4.1e-07; Indels 74; Gaps 18;

Matches 104; Conservative 56; Mismatches 174; Indels 74; Gaps 18;

QY 87 SLSSGPL-LMELAFLLRHVSGVWITNORSQTDVYSLSEHMLNHGVOLPARGOE 145  
 DB 50 SFGGVATILSLVPLRLRSIGTEARFVIEBPTEFFVYVTKTHNL-----QGNB 99  
 QY 146 AVDIAMKADVIYN-TVAGKMDPVLKDHVETKPKILMWHKMGHYFVYVYKLPF 204  
 DB 100 SLGLTEEMKEVILVNRNSKFLDSSFDV-----LVHDPQALILEFEYKKSFP 150  
 QY 205 VAGAMID-SHTTAEVYN-----SRTSDRLKIQPOTVYVHLGNSKELM-----EVAEDN 252



Db 151 LMRCHIDLSGPNRPFWRFEKRYRIFHLPEYVQPELDNRKNAIMPSSIDPLSEKN 210  
Qy 253 VARR---VLRHRIRESIGRSDLLFAINSYR---GCGDQLQAFYQALDLOHEKL 306  
Db 211 VELKQTELR--ILERDVPPEK---PITTVSSFPDMWG-----IFVITETRYKKE 258  
Qy 307 KVPRIHAIVVG---SDVNAQTFETQLR---DPVVK---NTIHRVHFVNKTAAVAPY 355  
Db 259 KIPGVQLLVGVMAHDPEGMIFYPEKTLRKIGEDYDVKVTNLIGVHAREVN-----AF 312  
Qy 356 LMAIVLVNQGSGEGFGRITTEMAFKLPIVGTAGGTTEIIVDGSSTGLHPAGKEGV 415  
Db 313 QGRSDVILQMSIREG--FGLTTEMMKQKVIQRAVGKIQYIDGTEGIVRDAENAV 370  
Qy 416 APLAKNIYRLASHAEOVSNGEKYGRYKEMFEMHMAERIAVLKDV 463  
Db 371 ----EKVLYLKHPVSKWGAKEKAKERYKKNFITTKEHMERYDIINSL 414

## RESULT 8

Q9WZ90 PRELIMINARY; PRT; 388 AA.  
AC Q9WZ90;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE Lipopolysaccharide biosynthesis protein, putative.  
GN TM0622.  
OS Thermotoga maritima.  
OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.  
OX NCBI\_TaxID=2336;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MSB8 / DSM 3109;  
RX MEDLINE=99287316; PubMed=10360571;  
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,  
Hate D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,  
McDonald L., Ueberlack T.R., Malek J.A., Linher K.D., Garrett M.M.,  
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,  
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,  
R. Salberg S.L., Smith H.O., Venter J.C., Fraser C.M.;  
RT "Evidence for lateral gene transfer between Archaea and Bacteria from  
genome sequence of Thermotoga maritima."  
RL Nature 399:323-329(1999).  
DR EMBL; AE001736; MADS5706.1; -  
DR TIGR; TM0622; -  
DR InterPro; IPR001296; Glycos\_transf\_1.  
DR Pfam; PF00534; Glycos\_transf\_1; 1.  
KW Complete proteome.  
SQ SEQUENCE 388 AA; 43648 MW; 6BA1BE01157F306 CRC64;

Query Match 7.8%; Score 187.5; DB 16; Length 388;

Best Local Similarity 23.5%; Pred. No. 2e-06;

Matches 88; Conservative 58; Mismatches 136; Indels 93; Gaps 13;

Qy 136 VQVLPARGQAVDIALKADVLITNTAVAGKMDLPVLKQHVKLPKIMWIMHMGCHFK 195  
Db 47 VEKLTSGQYVYTIYDYE-----AIAPS-----KVIRLLRAIKMKRRTNL 89  
Qy 196 VEYVK-----HLFVAGAMIDS-----HTTAE-----YNSKTSIDRLKI 229  
Db 90 LREIRPIIHSHLSARIALIPALLCRIPKVTHTITVAEKDAGKTRFPN-----RI 142  
Qy 230 QMPQTYVHLGNSKELMEVAEDNARVLAENHRESLGV-----RSEDLLFAI 277  
Db 143 AKFEGFVPVSIQEVASVKKLVGRKISTPVIYNGIDQKFSIDQKRDRLKITILNV 202  
Qy 278 INSVSRGKQDLELQAFYQALQLOHEKLVKPRIHAIVVGVSDVNAQTFETQLRDFVKN 337  
Db 203 AR-LSREKNALLVRAFSKAVQ-----SCPMLELMLVGDG-----ELRDLIELVQOL 249  
Qy 338 TIHSDVHFVNKTAAVAPYLAIDVLVNSQSGEGFGRITTEMAFKLPIVGTAGGTTE 397

Db 250 GLEEKVFEGVSDVPELLSQADIFVLSDYEG--FGLVAMAMAGLPVITATIGIPE 307  
Qy 398 IVDGSTGLHPAGKEGVAPLAKNIYRLASHAEOVSNGEKYGRYKEMFEMHMAERIA 457  
Db 308 ILEGGRAGLIVP--PROVDALAKAIVELARDEKRAELSDVGRULVAERF----- 355  
Qy 458 AVLKDVLRKSGEHSR 472  
Db 356 ----DIRRTVREYER 366

## RESULT 9

Q95SA2 PRELIMINARY; PRT; 424 AA.  
AC Q95SA2;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)  
DE GM04690P.  
GN CG1291.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
Gonzalez M., Guarini H., Li P., Liao G., Miranda C.J.,  
R. Munoz J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
Yu C., Lewis S.E., Rubin G.M., Ceiniker S.;  
RL Submitted (JOCF-2001) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AY060893; AAL28441.1; -  
DR FlyBase; FBgn0035401; CG1291.  
DR InterPro; IPR001296; Glycos\_transf\_1.  
DR Pfam; PF00534; Glycos\_transf\_1; 1.  
SQ SEQUENCE 424 AA; 48639 MW; F39F08BF7A11B35 CRC64;

Query Match 7.6%; Score 182.5; DB 5; Length 424;

Best Local Similarity 22.2%; Pred. No. 5.8e-06;

Matches 101; Conservative 75; Mismatches 179; Indels 99; Gaps 15;

Qy 79 VLVSHSLSGGFLMLMELAFLLRHVSQVWVITNRS-----QETNDVYSLEHRMLN 133  
Db 4 VLFHPDLGIGABRLVVDAAALKEKRGHVSFLTNHDSHCFEKLADGTFPV----- 57  
Qy 134 HGVQVLPARGQAVDIALKADVLITNTAVAGKMDLPVLKQHVKLPKIMWI 186  
Db 58 HVGQDLPARGLGRFPAICAVLMLYAAIYASFMPQREQDVVVCGLISVCP-----V 112  
Qy 187 HEKRGHVFYVYVHLPFVAGAMIDSHTTAEVNSRSDKIQMPQTYVHLG-----N 241  
Db 113 LRAPRPFPVLFYCHFP-----DQLSSRBGLKRLYRLPINMLEHTTGLADKVLVN 165  
Qy 242 SKELMEVAEDNAR-----RVLR-----EHRESIGVRSIDL-----LF 275  
Db 166 SKFTLAVPDQTRRLSTVPDVLVYSLHTGYFPOKQKLEORSALLDEPVHPRVLANFYT 225  
Qy 276 AINSVSRGKQDLELQAFYQALQLOHEKLVKPRIHAIVVGVSDVNAQTFETQLRDFVKN 332  
Db 226 LDINRVERKKHALLHSLRLGLDMLPATERK--RCRLIIAGVDTRCMENVEHFALHEH 283  
Qy 323 QTFKETQLDPAV-----KNTIHRVHFVNKTAAVAPYLAIDVLVNSQSGEGFGRIT 377  
Db 284 LTE-ELKLDQHVVALRSPTDEKCRLLFPAHCLLYTP-----ENEHFGIVP 328  
Qy 378 IEAMAFPLVLTAGGTTEIIVDGSSTGLHPAGKEGVAPLA\* IYRLASHAEOVSNGE 437  
Db 329 LEGNYCKPVLVNSGGTTEIVVNTSTGFLCEKTEKS---F\*AMQLPRDQLRVKMGD 385  
Qy 438 KGYGRVEMFEMHMAERIAVLKDVLRKSGEHS 471

Db 386 QGHKRVQKESFQAFADRLNGIIRDLVPISRESS 419

# RESULT 10

ID 059512 PRELIMINARY; PRT; 381 AA.

AC 059512; 01-AUG-1998 (TReMBLrel. 07, Created)

DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)

DE 01-MAR-2002 (TReMBLrel. 20, Last annotation update)

OC Pyrococcus

OS Pyrococcus horikoshii

NCBI\_TaxID=53953;

RP SEQUENCE FROM N.A.

RC STRAIN=OT3;

RA MEDLINE=98344137; PubMed=9679194;

RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,

RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,

RA Sakai M., Oguni K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,

RA Furuhashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,

RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,

RA Masuho Y., Shizuya H., Kikuchi H.,

RT "Complete sequence and gene organization of the genome of a hyper-

RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."

RL DNA Res. 5:55-76(1998).

DR EMBL; AF000007; BAA30965.1; -

DR InterPro; IPR001296; Glycos\_transf\_1.

DR Pfam; PF00534; Glycos\_transf\_1; 1.

DR Hypothetical protein; Complete proteome.

SW SEQUENCE 381 AA; 42826 MW; EDOBOCAQCFSCAEI CRC64;

Query Match 7.6%; Score 181.5; DB 17; Length 381;

Best Local Similarity 23.1%; Pred. No. 5.9e-06;

Matches 93; Conservative 67; Mismatches 161; Indels 81; Gaps 15;

QY 98 LAFLRHVGSQVWVITNQSGETNDVYSLHRLNMGVQLPARG--GEANDIALKADL 155  
 DB 27 LAIKLRERHGVGIYVNNRPT-----GSEELKRYGIDELIKIPGISPLDVLTYGL 79  
 QY 156 VILNTAVAGKMLDPVLDKDVHPKVLKILMIHMGHYFKVEYVHLR--FVAGAMID 211  
 DB 80 -----KSEELNEFLKD-----FDIHS--HHAFTPLSLKALKAGKMKETILT 122  
 QY 212 SHTTAEYMSRTSDRLKIOMP--QTYVH-----LGSKELMEVADNVARL----- 258  
 DB 123 THSISFAHESKLDWTLGFTIPLFKSYLAKSHIIVASQASPIEHFTSVPLVIVNGVD 182  
 QY 259 -----REHRESLGVSEDLFAIINSVSRKGQDLFLQAFYQALQIÖHEKLV 309  
 DB 183 DERFPARDKEKIKAKFGLGNVLY--VSRYSYRKGPHVILNAF-----SKIE 229  
 QY 310 RIHAVVVGSD-----VNAQKFEQTLRDFVKNITLHDVHVKNKTLAVALPYLAIDVYL 354  
 DB 230 DATLVMGNGEMLPFLKQTKF-----LGINKVYFNGVYPPDILPEYFRAVDYFVPS 283  
 QY 365 NSQGRGCGFRITIAAFKLPVGLGAGGTEIYLDGSGLLHPAGKEGVAFLAKNIV 424  
 DB 284 IS--SEAFGIYILEAMSGVPIIADVGIEVIVKESAGLLVPPGNE--LKLREAIK 338  
 QY 425 LASHAEQVYSGEKGQYKVEFMHMAERITAAVLKQVLRK 466  
 DB 339 LAKKEELRKVYGNNGRSVEEKYSNKKIVYKIERIYNVÖE 380

# RESULT 11

ID 09HH00 PRELIMINARY; PRT; 412 AA.

AC 09HH00; 01-MAR-2001 (TReMBLrel. 16, Created)

DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)

DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE Putative trehalose synthase.

GN PF1742.

OS Pyrococcus furiosus, and

OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;

NCBI\_TaxID=2261, 2265;

RP SEQUENCE FROM N.A.

RC SPECIES=P. furiosus, and T. littoralis;

RA MEDLINE=2056786; PubMed=1115105;

RA Dittuggero J., Dunn D., Maeder D.L., Holley-Shanks R., Chataud J.,

RA Horlacher R., Robb F.T., Boos W., Weiss R.B.;

RT "Evidence of recent lateral gene transfer among hyperthermophilic

RT archaea."

RL Mol. Microbiol. 38:684-693(2000).

DR EMBL; AE010272; ALB1866.1; -

DR EMBL; AF307053; AAG45391.1; -

DR EMBL; AF307052; AAG45375.1; -

DR InterPro; IPR001296; Glycos\_transf\_1.

DR Pfam; PF00534; Glycos\_transf\_1; 1.

DR Hypothetical protein; Complete proteome.

SW SEQUENCE 412 AA; 48017 MW; 6A5338298A6B5CBA CRC64;

Query Match 7.5%; Score 180; DB 17; Length 412;

Best Local Similarity 23.3%; Pred. No. 8.7e-06;

Matches 99; Conservative 55; Mismatches 158; Indels 112; Gaps 18;

QY 87 SLSGPL-LIMELAFLRHVGSQVWVITNQSGETNDVYSLHRLNMGVQLPARG--GEANDIALKADL 155  
 DB 48 SFGGVAERLHNLVPLMRDVGIDARWVLEGTFEFTVTKSF-HNALQGNKELRLTEEMK 106  
 QY 139 ---LPARGQEAVIDAL-KADVLNTAVAGKMLDPVLDKDVHPKVLKILMIHMGHYFKVEYVHLR--FVAGAMID 211  
 DB 107 KLYLEINKKNAEDIDLQPPYVLI-----HDPPAPLIEF----- 141  
 QY 195 KIVYVHLRPFVAGAMID-SHTTAEYMSRTSDRLKIOMPQTYVHVLGN----- 241  
 DB 142 ---YERQEWLRCHIDLSDPNLEFVKL--RQFEKDYRIFPMEEYVQEDLNQKVV 195  
 QY 242 -----SKELMEVADNVARVLRHRESLGVSEDLFAIINSVSR--GKGQDL 289  
 DB 196 IMPPSIDPLSEKNEMLSESELKTLERDV-----DPRPIITGVAFDPWKQ--- 243  
 QY 290 FLQAFYQALQIÖHEKLVKPRIHAVVCG---SPVNAQKFEQTLR--DFVYK----- 336  
 DB 244 ---VFVDIVYRKVSKLPEVQDLVGVMHADDPGGMIFPKTLRKIGEDYDIKYLITNL 299  
 QY 337 NTHDRVHFVKNKTLAVALPYLAIDVLYONSQGRGCGFRITIAAFKLPVGLGAGGTEIYLDGSGLLHPAGKEGVAFLAKNIV 424  
 DB 300 TGVHAR-----EVNAFQASDVITLQMSIRG--FGLTVEAMMKKEKPYVRAVGGLK 349  
 QY 397 EIVLDGSGLLHPAGKEGVAFLAKNIVLASHAEQVYSGEKGQYKVEFMHMAERITAAVLKQVLRK 466  
 DB 350 LQIVDQKGFPL---VKDVNDALIKTYLLBHHQVQAEQKNAEKRIKENPITTKLERY 405  
 QY 457 AAVL 460  
 DB 406 LDLL 409

# RESULT 12

ID 09VZ08 PRELIMINARY; PRT; 424 AA.

AC 09VZ08; 09VZ08;

DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
 DE CG1291 protein (RE67594p).  
 GN CG1291  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 [1]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Rasmussen M.D., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,  
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 Abril J.F., Aghayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,  
 Baller R.M., Basu A., Bakendale U., Bayraktaroglu L., Beasley E.M.,  
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 Bokkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 Foeller C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glasner K.,  
 Glodek A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,  
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 Svirskas R., Tector C., Turner R., Venter E.H., Wang A.H., Wang X.,  
 Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weisenbach J.,  
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhou X., Zhao Q., Zheng L.,  
 Zheng X.H., Myers E.W., Rubin G.M., Venter J.C.;  
 "The genome sequence of *Drosophila melanogaster*.";  
 Science 287:2185-2195(2000).  
 [2]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN=BERKELEY;  
 RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,  
 Champe M., Chavez C., Dorsett V., Dresnak D., Fatfin D., Fiske E.,  
 Georgi R., Gonzalez M., Guatin H., Krommiller B., Li P., Liao G.,  
 Miranda A., Mungall C.J., Nuccio J., Pacle J., Pargias V., Park S.,  
 Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 Ceiniker S.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: AEO03477; AAF47219.1; -;  
 DR EMBL: AY071629; AAL49251.1; -;  
 DR FLYBase: FBgn0035401; CG1291.  
 DR InterPro: IPR001296; Glycos transf. 1.  
 DR Pfam: PF00534; Glycos transf. 1; 1.  
 SQ SEQUENCE 424 AA; 48598 MW; 9F77DDA4A610DAB34 CRC64;

Query Match 7.5%; Score 179.5; DB 5; Length 424;  
 Best Local Similarity 22.2%; Pred. No. 9.9e-06;  
 Matches 101; Conservative 73; Mismatches 181; Indels 99; Gaps 15;

QY 79 VLVVSHLSLSGGPRLIMELAFLLRHVGSQVWITNORS-----QETNDVYSLBHRMLN 133  
 DB 4 VLFPLPDLGIGABERLVVDALALKEKGHVSLTNHDSHCKETADGTPV----- 57  
 QY 134 HGVQVLARQGEVDIALKADLVLTNTAVAGK-----LDPLKDHVKVLPKILWMI 186  
 DB 58 HVGDMPLPRLLFGFRFYALCATLRVLAITAYASFMPQEQDVVVCDSVCP-----V 112  
 QY 187 HEMGHVFKVEYVKLPFVAGAMIDSHTTAEWNSRTSDRLKIMQPTVYVHLG-----N 241  
 DB 113 LRFAHPRKVLFCYCHP-----DQLSSREGLLKRLYLRLPIWMEHTIGLADKVLVN 165  
 QY 242 SKELMEVAEDNVAR-----RLR-----EHIRESCVRSBDL-----LF 275  
 DB 166 SEFTLRVQDTEFRSTVDDVLPSTHTQYFDQMDKLEQSRSSALDEPVHPRVPLNATY 225  
 QY 276 ALINSVSRGGQDLFLQAFQYALQIQHEKTKAPRIHVVVG-----SDVNA 322  
 DB 226 LDIVRYERKKNHALHLSRLILGMLPATDFK--KRLILINGYDTRCMENVEHFALEH 283  
 QY 323 QTKPEQTQDLPFV-----KNTHRHVFNPKTLAVALPILADLVVNSGSGEGEGRIT 377  
 DB 284 LTF--ELKIQHVVLLSPDDEKCRLLFAHCLLYTP-----ENEHGIVP 328  
 QY 378 IEAFAFKLPVLTAGCTTEIVLDGSTGLLHPAGKEGVAPLAKNIVLASHAEQVNSGE 437  
 DB 329 IEGWCSKPVVALNSGPTETVSTGFLCEKTEKS--FGAMQLQFPDQQLRVKMGD 385  
 QY 438 KGYGRVEMEMEHMARIANLVDRKQSHS 471  
 DB 386 QGHRKVOOKFSFOAFADRLNGILRDVLPSIRESS 419  
 RESULT 13  
 ID OPEZ9 PRELIMINARY; PRT; 370 AA.  
 AC OPEZ9;  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DE 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
 DR LipoPolysaccharide biosynthesis protein.  
 GN XF0879.  
 OS Xylella fastidiosa.  
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;  
 CC Xylella.  
 NCBI\_TaxID=2371;  
 [1]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN=9A5C;  
 RX MEDLINE=20365717; PubMed=10910347;  
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,  
 Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,  
 Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,  
 Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carer H.,  
 Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,  
 Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,  
 Facichiani A.P., Ferreira A.J.S., Ferreira V.C.A., Fetto J.A.,  
 Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan J.R.,  
 Garmier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,  
 Ho P.L., Hobeisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,  
 Krieger J.E., Kuranae E.E., Laigret F., Lambais M.R., Leite L.C.C.,  
 Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,  
 Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Matucuma C.L.,  
 Marques M.V., Martins E.A.L., Martins E.M.F., Matucuma A.Y.,  
 Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,  
 Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,  
 RA Nani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,  
 de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,  
 RA Paixoto B.R., Pereira G.A.G., Pereira H.A.Jr., Pasquero J.B.,  
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,  
 de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawaaki H.E.,  
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,  
 da Silveira U.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,

Matches	16;	conservative	53;	mismatches	124;	indels	45;	gaps	14;
QY	160	TAV-AGKWLDPVLKDHVPKVLKILLMIHMRGHYFKVEYVK-HLPFVAGMIDSHITLAE	217						

[illegible]

Tue Jul 8 09:50:08 2003

us-09-938-294-45.rpt

Page 9

Db	308	KAGAMELVBHGVNGFLTPGSEOLANIINTCIEDTOKTATIASMAQAIASQRPVVYI	367
Oy	436	GEK 438	
	:	:	
Db	368	NOQ 370	

Search completed: June 30, 2003, 16:27:36  
Job time : 41 secs

---

